Populacijsko-genetička struktura čovječje ribice (Proteus anguinus) u jamskom sustavu Postojna-Planina

Kermek, Dora

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:568804

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2025-02-22**

Repository / Repozitorij:

Repository of the Faculty of Science - University of Zagreb

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Dora Kermek

Populacijsko-genetička struktura čovječje ribice (*Proteus anguinus*) u jamskom sustavu Postojna-Planina

Diplomski rad

Zagreb, 2024.

University of Zagreb

Faculty of Science
Department of Biology

Dora Kermek

Population genetic structure of the olm (*Proteus anguinus*) in Postojna-Planina Cave System

Master thesis

Zagreb, 2024.

Ovaj rad je izrađen u Laboratoriju za podzemnu biologiju Oddelka za biologijo Biotehničkog fakulteta Sveučilišta u Ljubljani, pod mentorstvom doc. dr. sc. Valerije Zakšek, te komentorstvom izv. prof. dr. sc. Tvrtka Dražine. Rad je predan na ocjenu Biološkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu radi stjecanja zvanja sveučilišna magistra molekularne biologije.

Najveću zahvalu upućujem mentorici doc. dr. sc. Valeriji Zakšek. Za sve savjete i rasprave prilikom izrade i pisanja rada, pruženu podršku i razumijevanje, strpljenju, te ponajviše na ukazanom povjerenju. Hvala komentoru prof. dr. sc. Tvrtku Dražini na korisnim savjetima i što je bio uvijek dostupan za sve moje nedoumice.

Hvala prof. dr. sc. Peteru Trontelju što mi je pružio priliku da se pridružim radu laboratorija, te na dragocijenim diskusijama i pomoći tijekom izrade rada. Hvala svim članovima SubBioLab-a što su mi učinili boravak u Sloveniji ugodnim i vrijednim povratka.

Hvala mojim biospelićima iz sekcije, uz vas sam se zaljubila u terene i biospeleologiju. Hvala članovima Hrvatskog biospeleološkog društva što su me prigrlili u svoj svijet, pružili brojne znanstvene prilike i otvarali vrata za različita istraživanja. Bez vas ne bi bilo ovog diplomskog rada.

Posljednju zahvalu upućujem svojoj obitelji, mami i tati koji me od samog početka bodre u svim prilikama i neprilikama u kojima se nađem. Za vas nijedan problem nije neriješiv, a mislim da je to najveća pouka koju sam mogla izvući. I možda da treba malo više paziti na sebe. Puse!

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek Diplomski rad

Populacijsko-genetička struktura čovječje ribice (*Proteus anguinus*) u jamskom sustavu Postojna-Planina

Dora Kermek

Rooseveltov trg 6, 10000 Zagreb, Hrvatska

Čovječja ribica (*Proteus anguinus*), prepoznata kao prva opisana špiljska životinja, ujedno se smatra simbolom očuvanja podzemnih ekosustava Dinarskog krša. Sa svojom velikom i relativno lako dostupnom populacijom čovječje ribice, jedan od najvećih jamskih sustava u Sloveniji, jamski sustav Postojna-Planina, predstavlja odličnu priliku za proučavanje karakteristika populacije pomoću populacijske genetike. Korištenjem skupa specifičnih i polimorfnih mikrosatelitnih biljega, zajedno s opsežnom zbirkom briseva kože, proučava je populacijska struktura čovječje ribice. Unutar jamskog sustava populacije Postojnske jame i Planinske jame genetski su odvojene. Određeni su populacijsko-genetički parametri obiju populacija. Metodom ponovnog ulova uhvaćenih jedinki izvršenoj 2015. i 2016. godine dobiven je omjer ponovnog hvatanja od 14.1%. Većina ponovno uhvaćenih jedinki pronađena je na istom dijelu Planinske jame kao i tijekom prvog ulova, indicirajući visoku vjernost istom području kod jedinki proučavanih populacija. Ovaj rad predstavlja prvi pokušaj razjašnjavanja srodstvenih odnosa među čovječjim ribicama. Analiza genotipa pokazala se dostatnim za određivanje braće i sestara te polubraće i polusestara unutar istraživanih dijelova sustava.

Ključne riječi: biologija podzemlja, konzervacijska genetika, ponovni ulov, analize srodstva
(56 stranica, 17 slika, 6 tablica, 75 literaturnih navoda, jezik izvornika: engleski)
Rad je pohranjen u Središnjoj biološkoj knjižnici

Mentor: doc. dr. sc. Valerija Zakšek
Komentor: izv. prof. dr. sc. Tvrtko Dražina

Ocjenitelji:
izv. prof. dr. sc. Tvrtko Dražina
izv. prof. dr. sc. Ivana Buj
izv. prof. dr. sc. Dubravko Pavoković

Rad prihvaćen: 5. rujan 2024.

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Science
Department of Biology

Master thesis

Population genetic structure of the olm (*Proteus anguinus*) in Postojna-Planina Cave System

Dora Kermek

Rooseveltov trg 6, 10000 Zagreb, Croatia

The olm (*Proteus anguinus*), recognized as the first described subterranean animal, stands as a symbol for the conservation of Dinaric karst underground ecosystems. One of the biggest cave systems in Slovenia, the Postojna-Planina Cave System, with its large and relatively easily accessible olm population at various sites, provides an excellent opportunity to study population characteristics using population genetics. By utilizing a set of specific and polymorphic microsatellite markers along with the high number of skin swab samples, detailed population structure of olms has been studied. Within the Cave System, the olm populations from Postojna and Planina Cave are genetically separated. Population genetic parameters of both populations were determined. The recapture study from 2015 and 2016 resulted with a recapture ratio of 14.1%. Most of the recaptured individuals were found on the same section of the Planina cave as first captured, indicating very high site fidelity in the studied population. This study is the first insight into the relatedness and parentage of olms. The analyses showed potential for the determination of full and half-sibling relationships based on available genotypes.

Keywords: subterranean biology, conservation genetics, recapture, parentage analysis
(56 pages, 17 figures, 6 tables, 75 references, original in: English)
Thesis is deposited in Central Biological Library.

Mentor: Prof. Valerija Zakšek, PhD
Co-mentor: Assoc. Prof. Tvrtko Dražina, PhD

Reviewers:
Assoc. Prof. Tvrtko Dražina, PhD
Assoc. Prof. Ivana Buj, PhD
Assoc. Prof. Dubravko Pavoković, PhD

Thesis accepted: September 5, 2024

Table of Contents

ABBREVIATIONS

ADO – allelic dropout

AMOVA – analysis of molecular variance

bp – base pairs

DyadML – dyadic likelihood estimator

f – inbreeding coefficient, calculated by inbreeding estimators

FA – false allele

F_{IS} – inbreeding coefficient, calculated by *F*-statistics

FL – full-pedigree likelihood

FS – full-sibling

F_{ST} – fixation index

H_e – expected heterozygosity

H_o – observed heterozygosity

HS – half-sibling

IBD – identical by descent

IUCN – International Union for Conservation of Nature

M – median

MCMC - Markov chain Monte Carlo

ML – maximum likelihood

MRR – mark-release-recapture

mtDNA – mitochondrial DNA

N_a – number of detected alleles for each locus

N_e – effective number of alleles for each locus

Q_1 – first quartile

Q_3 – third quartile

PO – parent-offspring

PPCS – Postojna-Planina Cave System

R – allelic richness

T_a – annealing temperature

TrioML – triadic likelihood estimator

1. INTRODUCTION

1.1. Karst

Karst represents a specific type of diverse terrain characterized by high rock solubility and porose underground (Gunn 2004; Jones and White 2019). It covers 15.2% of the worldwide ice-free land areas and 21.6% of European land surfaces (Chen et al. 2017; Goldscheider et al. 2020). A necessary factor for the formation of the karst is water and landscape hydrology. Dissolution of the rocks, mostly carbonate rocks (limestone and dolomite), enables circulation of the underground water instead of flowing along the surface in river channels, therefore distinguishing karst terrain from fluvial, coastal, glacial, and other systems (Gunn 2004). The karst ecosystem encompasses the entire land area and aquifer volume responsible for draining water, covering both surface and subterranean habitats (Jones and White 2019).

1.2. Subterranean habitats and subterranean biodiversity

Subterranean habitats are characterized by specific abiotic and biotic features: absence of light, limited changes in temperature, food scarcity, and high physical fragmentation. Terrestrial subterranean habitats include the entire vadose (unsaturated) zone of the underground, particularly evident in karstic regions, such as caves, fissures, and cracks. Aquatic subterranean habitats include any water stored underground, encompassing 97% of all unfrozen freshwater on the planet (Gibert and Deharveng 2002). Considering their prevalence, subterranean ecosystems are believed to be the most widespread nonmarine ecosystems on the planet (Mammola et al. 2019). Despite specific and unwelcoming features, subterranean habitats host a wide range of specialised organisms. However, studies of the subterranean biodiversity encounter various difficulties, stemming from the challenges in accessing the habitats, understudied species diversity, to the predominance of β-diversity over α-diversity, and cryptic diversity (Deharveng et al. 2024). Based on the current estimates, most of the obligate subterranean organisms are still undiscovered and unknown for science (Zagmajster et al. 2018).

Groundwater, as underground water kept in the captivities of consolidated and unconsolidated rocks, hosts various aquatic subterranean organisms. It serves as a vital reservoir of drinking freshwater, playing a critical role in human sustenance. An estimated 25% of the world's population depends wholly or partially on drinking water drawn from groundwater aquifers (Mammola et al. 2019). Strong interactions between surface aquatic ecosystems and

groundwater ecosystems lead to enhanced deterioration of groundwater biodiversity, endangering subterranean organisms (Kretschmer et al. 2023; Malard et al. 2023). The most challenging threats impacting subterranean biodiversity worldwide are habitat loss, groundwater overexploitation and contamination, climate change, and intrinsic vulnerability of the subterranean fauna (Mammola et al. 2019). Conservation efforts often depend heavily on appealing and charismatic species, termed as *flagship species*, that garner public support and secure funding from the broader community. The European cave salamander (*Proteus anguinus*) is recognized as the groundwater flagship species, with a central role in the conservation of the Dinaric Karst caves and subterranean biodiversity (Kostanjšek et al. 2023).

1.3. The olm (*Proteus anguinus*)

The European cave salamander, *Proteus anguinus* Laurenti 1768, also known as olm or simply proteus, is the first described cave organism and the only exclusively cave-dwelling chordate species in Europe (Figure 1.). It is endemic to Dinaric Karst, with a distribution ranging from the Isonzo-Soča River in southeastern Venezia Giulia in Italy to the Trebišnjica River in eastern Herzegovina (Sket 1997). Although mainly inhabiting subterranean waters, the olms are regularly present in some springs. During periods of high rainfall and floods, the olms can also be found in rivers connected to the underground flow (Manenti et al. 2024). Serving as a flagship species in the conservation efforts of the Dinaric karst underground, it indirectly influences the preservation of cave ecosystems and subterranean species. It is protected as a priority species by the European Habitats Directive (92/43/EEZ) and the IUCN Red List listed as a vulnerable (VU). Italy, Slovenia, Croatia, and Bosnia and Hercegovina protect proteus on a national level (Zakšek and Trontelj 2017; Zakšek et al. 2018).

1.3.1. Morphology

Proteus' body is elongated, with a total length of 25 – 35 cm (Koller Šarić et al. 2019). Legs are short compared to the body, having three toes on the forelimbs and two toes on the hindlimb. The tail is also short, flattened, and fin shaped. Its skin lacks pigmentation; hence the body colour is white to pinkish due to the blood capillary positioned near the skin's surface. Three pairs of red gills, together with lungs and skin, are used for breathing. The eyes are delineated on the skin's surface but are overlaid by an additional skin layer. Given description refers to the white morphotype of proteus (*Proteus anguinus anguinus* Laurenti 1768) (Figure 1), however current taxonomy recognizes one more subspecies of the olm: *Proteus anguinus parkelj* Sket and Arntzen 1994. *P. a. parkelj* is pigmentated, has functional eyes, and differs from the white

morphotype also in the skull morphology. Although lacking troglomorphic features, dark proteus is considered a troglobiont, rarely appearing outside the caves (Sket 2017).

Figure 1. The adult of olm (*Proteus anguinus anguinus*) from Planina Cave. Photographed by: Valter Leban.

1.3.2. Reproductive biology

Neoteny of proteus body causes the lack of sexual dimorphism, making external morphological sex identification difficult. Little is known about olms reproduction in natural environments, but scientific observations of animals in captivity offer valuable insights into this aspect of proteus biology. Oviparity in proteus was confirmed multiple times (Koller Šarić et al. 2019). Females become sexually mature at the age of 15, and lower temperatures favours prolongation of reaching this period. It is assumed that males reach maturity earlier (Juberthie et al. 1996). Both sexes mature having body length between 14 – 18 cm, however they start to reproduce when they reach 20 – 24 cm (Durand and Delay 1981). In captivity, females lay eggs every six years during the minimum 30-years reproductive period (Juberthie et al. 1996). During mating, males become significantly territorial, circling the female entering their territory. After the male deposits spermatophore, the female harvests them with cloaca and stores in the spermatheca. Subsequently, internal fertilisation takes place and the female lays the eggs outside the male

territory (Briegleb 1962). Females usually lay between 20 and 60 eggs (Juberthie et al. 1996). The length of embryonic development is between two and six months (Guillame et al. 1999).

1.3.3. Evolution and phylogenetic relationships

Olm belongs to the family of Proteidae, which comprises two extant genera of permanently aquatic and neotenic salamanders. The genus, *Necturus*, commonly known as mudpuppies, encompasses five currently known species inhabiting various types of surface waters in eastern North America (Gorički and Trontelj 2006). The genus *Proteus* in its current taxonomic perception comprises two subspecies of olm: *P. a. anguinus*, and *P. a. parkelj,* as described in the *Morphology* subsection. However, recent molecular analyses using large-scale mitochondrial DNA (mtDNA) sequence data in a combination with genome-wide single nucleotide polymorphisms (SNPs) propose division of proteus in nine species-level lineages (Recknagel et al. 2024b). According to Recknagel et al. (2024b) these lineages are named: Istra, Krajina, Para-Littoral, Lika, Kras/Carso, Ljubljanica, Dolenjska, Stična and Parkelj. Parkelj lineage represents morphologically distinct specimens of dark morphotype belonging to the previously described subspecies.

1.4. Population and conservation genetics

Population genetics is a branch of genetics studying genetic variation within and among present populations, evolutionary factors explaining this variation, as well as changes in alleles frequencies and genotypes over time. The foundations of population genetics and the main concept of particulate inheritance were laid by Gregor Mendel (19th century), upon which the concept allowing determination of the allele frequencies based on the genotype's frequencies (20th century, called Hardy-Weinberg equilibrium) enabled further understanding of the genetic structure of the population. Along with the Hardy-Weinberg equilibrium and the deviation from it, linkage disequilibrium represents an important population characteristic in genetic analyses (Hamilton 2009). Linkage disequilibrium (LD) or gametic disequilibrium is the non-random correlation of alleles across various loci (Slatkin 2008). Linkage disequilibrium occurs because some loci are located on the same chromosome, making them physically linked. The distribution of loci in segregated cells after meiosis will be affected by the recombination rate between sister chromatids (Hamilton 2009). Although detecting LD does not ensure either linkage or the absence of equilibrium, it reflects the population history, the breeding system, the pattern of geographic subdivision, natural selection, gene conversion, and mutations (Slatkin 2008).

Conservation genetics appears simultaneously with the emergence of conservation biology, with the aim to preserve species by applying genetic methods. Theoretical roots of conservation genetics lie in population genetics, while its application in wildlife preservation enabled the rapid growth of the field (Willi et al. 2022). Main concerns in wildlife conservation include the identification of populations and conservation units, detection of hybridization, estimations of inbreeding, appraisal of population size, evaluation population's capacity to endure and adapt to environmental changes, along with with understanding the factors influencing this capacity. Information contained in genetic data can address all of the mentioned concerns, providing critical insights for wildlife management (Hohenlohe et al. 2021).

1.4.1. Molecular markers

A key element in population and conservation genetics research is the possibility to identify different genotypes. Due to this intrinsic potential, molecular markers, such as DNA fragments correlated with the specific location in the genome, represent a highly valuable tool in population studies. The most commonly used molecular markers are amplified fragment length polymorphisms (AFLPs), microsatellites, and single nucleotide polymorphisms (SNPs). AFLPs are dominant and nowadays outdated markers, whereas microsatellite and SNPs represent codominant and widely used markers. Microsatellites are highly polymorphic and offer relatively high statistical power for each locus. Microsatellites are short sequence repeats (SSRs) or tandem repeats consisted of di-, tri-, or tetranucleotide units (1-6 bps) which are interspersed in both coding and non-coding regions of the genome (Pathak and Ali 2012). The strand slippage during DNA replication is believed to be the main cause of microsatellite mutation, causing variations in the number of repeat unit(s). Consequently, the "stepwise mutation" model, in which alleles arise dependently of the previous allele and each mutational event results in the gain or loss of the single loci, is considered to be an adequate theoretical model of their evolution. Therefore, microsatellites are useful for population genetics studies, for example studying hybridization, inbreeding, genetic diversity and connectivity of populations, conservation biology, evolutionary history, and especially for analyses of parentage identification (Putman and Carbone 2014). However, they are affected by null alleles, homoplasy, and intricate, variable mutation processes that complicate the results (Hauser et al. 2021).

SNPs are biallelic and follow a less complex mutations model, nonetheless they are less informative per locus, requiring a greater number of SNP loci to achieve the same statistical power as microsatellites. SNPs are also affected by null alleles, but with their even distribution

across the genome, they provide better overview of genome-wide variations (Hauser et al. 2021). Comparative analysis of microsatellites and SNPs revealed that SNPs arose as better markers for estimation of genetic diversity and population structure (Morin et al. 2009; Muñoz et al. 2017). However, microsatellites performed equally or outperformed SNPs in parentage analysis (Weinman et al. 2015; Flanagan and Jones 2019)

The set of twenty-three novel polymorphic tetranucleotide microsatellite markers was developed and tested on olms (Zakšek et al. 2018), which enabled population genetics studies of olms, like genetic diversity and population size, as well as population structure and history of this subterranean species. Markers were developed and tested on the proteus population of the Postojna-Planina Cave System, enabling first insights into the population structure of olms. The first experiments started over a decade ago, which in the meantime led to the establishment of an extensive database of microsatellite genotypes across different regions of the Dinaric Karst. Markers also proved the potential for the assessment of effective population size through genetic mark-release-recapture method. However, no research has yet been conducted to examine their ability to determine relatedness and inbreeding among olms, as well as parentage. SNPs are also being used in proteus recently on smaller sample size (Recknagel et al. 2024a; Recknagel et al. 2024b).

1.4.2. Designating the population structure

Most species display measurable genetic differentiation among populations. The degree and pattern of this differentiation can vary significantly across species (Ward 2006). Furthermore, each population can express inner structure, and in many population studies, it is beneficial to understand the division of the population into subpopulations, population groups, or clusters. Individuals in the population can be assigned based on their genotypes by utilizing molecular markers, like microsatellites or SNPs (Pritchard et al. 2000). One of the most widely used methodologies in designating the population structure was introduced by Pritchard et al. (2000) and implemented into the software Structure. The model is based on Bayesian clustering approach with K populations, each characterized by a set of allele frequencies. The model assumes Hardy-Weinberg equilibrium and linkage equilibrium among marker loci within the population. Clustering methods used in the program are model-based methods. In model-based methods, standard statistical methods, like Bayesian methods, are used to designate the parameters of each cluster concurrently with designating the cluster membership of each specimen. Furthermore, two modelling assumptions are implemented: without admixture (each individual originates in a single population) and with admixture (individuals can inherit part of

their genome from ancestors of an unknown population). A characteristic feature of real genetic data is admixture among populations (Pritchard et al. 2000).

1.5. Relatedness

Relatedness among individuals indicates the presence of a shared recent common ancestor. Estimation of relatedness is based on the probability that a set of genes are identical-by-descent (IBD), meaning that they have been inherited from a single ancestral gene (Speed and Balding 2015). The coefficient of relatedness (*r*) measures the proportion of shared alleles between pairs of individuals that are IBD. Calculation of the relatedness between individuals within the population serves as a valuable tool for many genetic topics, including studies of gene flow, trait heritability, kin selection, cooperative breeding, social behaviour and structure, and in the management of conservation breeding programs (Taylor 2015).

Inbreeding refers to the mating of closely related individuals. The inbreeding coefficient (F_{IS} or *f*) measures the probability that two alleles at each analysed loci in an individual are IBD. Estimation of inbreeding coefficients is essential for studies of inbreeding depression. Both relatedness and inbreeding can be estimated directly from genetic markers and calculated at individual level, as well as averaged over populations (Taylor 2015). The inbreeding coefficient can be calculated using the *F*–statistics (F_{IS}) or inbreeding estimators (*f,* explained below).

Since the 1980s, seven commonly employed relatedness estimators have been developed, belonging to two different types: moment estimators and likelihood methods. Moment estimators use probabilities of identity by descent to calculate the relatedness between individuals (Queller and Goodnight 1989; Li et al. 1993; Ritland 1996; Lynch and Ritland 1999; Wang 2002), while likelihood methods estimate the probability of individuals belonging to a certain relationship given the available marker information (Anderson and Weir 2007; Wang 2007). From mentioned estimators, it is possible to calculate inbreeding coefficients using two moment estimators (Ritland 1996; Lynch and Ritland 1999) or both likelihood methods (Anderson and Weir 2007; Wang 2007). However, the performance of estimators is influenced by the relatedness structure of the studied population, the population's demographic background, the number of loci used, and their polymorphism. No single estimator for both relatedness and inbreeding excels across every scenario and it is advised to conduct *a priori* simulations in order to identify the most suitable estimator for a given scenario (Wang 2011; Taylor 2015). In recent literature and more recent studies, the mostly used estimator for the determination of both coefficients was the triadic likelihood method (TrioML; Wang 2007),

since this method proved to be the most accurate, having the lowest root mean error, exhibiting the lowest variance and highest correlation in on microsatellites (Wang 2007; Patenković et al. 2022; Lyu et al. 2023; Pacheco et al. 2024).

1.6. Parentage and sibship analysis

In studies of diverse ecological and evolutionary topics, parentage patterns play a central role. These include sexual selection, assessment of quantitative genitive traits, patterns of dispersal and recruitment, as well as wildlife management through conservation biology (Jones et al., 2010). Concepts in parentage analysis have remained largely unchanged over the past decade (Flanagan and Jones 2019). Essentially, parentage analyses compare the genotypes of candidate parents with the offsprings' genotypes with the aim of assigning possible families (Jones et al. 2010) The first review of parentage analysis was presented by Jones and Ardren (2003), with later supplementation and scrutinization by Jones et al. (2010) and Flanagan and Jones (2019). Based on the most recent review (Flanagan and Jones 2019), there are four main categories of parentage analysis: exclusion, parentage assignment, Bayesian parentage analysis, and parental and sibship reconstruction.

The exclusion method is based on the fact that each parent passes at least one allele per locus to each of its offspring. If such a match does not occur, the candidate parent is removed from the consideration (Jones and Ardren 2003; Jones et al. 2010).

The parentage assignment is divided into two categories: categorical and fractional allocations (Flanagan and Jones 2019). The categorical allocation designates the offspring to the candidate parent having the highest posterior probability or likelihood. The fractional allocation works similarly to categorical allocation; however, it allocates each offspring partially to each of the non-excluded candidate parents. Although without biological meaning, this designation has better attributes from the statistical point. Absolute likelihoods can be used to evaluate alternative hypotheses by calculating the likelihood ratio of one hypothesis against another, often a null hypothesis (Jones et al. 2010; Flanagan and Jones 2019).

Bayesian parentage analysis, also called full probability parentage analysis, evolved from a fractional allocation. The method estimates parent-offspring relationships with population-level variables, therefore including the uncertainty in parentage analysis with the estimation of variables of interest. Another advantage is the convenient integration of prior information, therefore incorporating information that implies even slight changes in the probability of parentage for particular individuals (Flanagan and Jones 2019).

Parental and sibship reconstructions are mostly based on maximum-likelihood approaches (Flanagan and Jones 2019). Parental reconstruction reconstructs parental genotypes by utilizing offspring genotypes from full- or half-sibling families, therefore relying on at least one shared parent among offsprings. Sibship reconstruction is used to reconstruct offsprings into different classes of relationships: full-siblings, half-siblings, and unrelated individuals. After the identification of full-sib and half-sib groups, genotypes of parents can be reconstructed (Jones et al. 2010).

The introduction of microsatellite markers, as the first widely available codominant, single locus, and polymorphic markers, marked the bloom of molecular parentage analyses. Ideally, parentage analyses should implement a high number of microsatellite loci with very high levels of polymorphism per locus (Jones et al., 2010).

Although microsatellites have enabled significant progress in parentage analysis, methodology encounters several problems. In species with an abundant number of highly polymorphic microsatellites, parentage analysis proved to be very successful, however many species manifest little polymorphism at microsatellite loci. Furthermore, the identification of microsatellite loci demands a large initial investment, requiring the design of locus-specific primers and optimization of PCR conditions. In addition, microsatellite-based study requires high labour resources and demands sparsely documented criteria for distinguishing true alleles from artefactual bands on sequencing gels (Flanagan & Jones, 2019; Jones & Ardren, 2003). Recently, SNPs are becoming a valuable alternative to microsatellites (Flanagan and Jones 2019).

The most important contributing factors determining the effectiveness of the parentage study are the sampling design and the choice of molecular markers. Successful parentage analysis necessitates either the use of highly polymorphic markers or a substantial number of markers with low to moderate polymorphism levels. Despite sampling design and choice of markers, inconsistencies in genotyping data cannot be omitted. Important groups of inconsistencies are genotyping errors, mutations, and null alleles. Genotyping errors arise when a genotype is misread, fails to amplify, or produces a misleading outcome. With the mutations, alleles inherited by the offspring differ from the ones occurring in the parent. Both errors result in incompatibilities between offspring and their true parent, with the possibility of eliminating the parent from consideration. Null alleles, also known as nonamplifying alleles, also cause an incongruence between parent and offspring. They are especially significant in heterozygotes with one amplified and the other nonamplified allele, therefore leading to mistaken recognition

of this heterozygote as a homozygote. Nonetheless, loci with null alleles can be identified as a departure from Hardy-Weinberg equilibrium at a given locus or as a non-Mendelian allele segregation in known family groups (Jones et al., 2010; Jones & Ardren, 2003).

Reviews from Jones et al. (2010) and Flanagan & Jones (2019) provide lists of available software programs, together with their characteristics and performances. Two widely used software programs for both parentage and sibship reconstruction are Colony 2.0.7.1. (Jones & Wang, 2010) and ML-Relate (Kalinowski et al., 2006), both having intuitive graphical user interface on the Windows operating system. ML-Relate implements the maximum likelihood method to estimate sibship and parentage for each dyad separately (Kalinowski et al. 2006). Colony implements full-pedigree likelihood estimates, which infers parentage and sibships jointly, while the likelihood is considered through the complete pedigree configuration, instead of only for dyads (pairs of individuals) (Jones and Wang 2010).

1.7. Postojna-Planina Cave System

Postojna-Planina Cave System (PPCS) is a system of six caves with separate entrances in Central Slovenia and connected by unknown underground passages. These caves are: Postojna Cave (slv. *Postojnska jama*), Planina Cave (slv. *Planinska jama*), Črna Cave (slv. *Črna jama*), Otok Cave (slv. *Otoška jama*), Pivka Cave (slv. *Pivka jama*), and Magdalena Cave (slv. *Magdalena jama*), possessing individual cadastre numbers in Slovene Cadastre of Caves (Zagmajster et al. 2021). These entrances are joined by the Lekinka and Tkalca Caves, connected with others through impassable flooded channels. The PPCS reaches a depth of 115 m, and its cumulative passage length surpasses 34 km, of which 24 km comes from Postojna Cave, and about 10 km is the sum of the length of Lekinka, Planina Cave, and Tkalca Cave (Šebela 2019; Zagmajster et al. 2021). The Pivka River sinks into the underground system of the Postojna Cave, in Planina Cave it joins with the Rak River (Rak Channel) and flows as the Unica River out of Planina Cave via Planinasko Polje. The cave system was developed in carbonate rocks from the Cretaceous period. Temperatures in the PPCS vary depending on cave entrances and distance from river sinks, ranging from 3 to 13°C, while inner parts have a constant temperature of 8°C. The Pivka River's temperature fluctuates daily and seasonally, with oxygen levels reflecting surface conditions near the sink and becoming saturated further downstream (Zagmajster et al. 2021). The discovery and scientific description of the cave-dwelling beetle from Postojna cave, *Leptodirus hochenwartii* Schmidt 1832, marks the

beginning of speleobiology. *L. hochenwartii* was the first scientifically described[1] subterranean animal in the world, making the Postojna-Plaina Cave System the "cradle of speleobiology" (Zagmajster et al. 2021).

PPCS is home to a large and relatively easily accessible population of olms. Approximately two kilometres of Pivka River inside Planina Cave is relatively easily accessible and therefore an appropriate site for the genetic study of the olm population. Based on mtDNA and SNPs, olms from PPCS belong to the Ljubljanica species-level lineage (Recknagel et al., 2024b). The first data obtained from PPCS microsatellite analysis revealed a weak genetic structure within this population (Zakšek et al. 2018), but the spatial distribution of individuals, migrations, and population size estimates in larger sample sizes remain unexplored.

1.8. Olm families and population structure in wildlife

Being the cave-dwelling organism, research of olm in wildlife encounters many obstacles. Access to their habitat is demanding, with the necessity to overcome technically demanding caves and pits. Most of the time, cave-diving methods are inevitable, and the number of observed and sampled individuals is small (Zakšek et al. 2018). Consequently, most of the findings about proteus ecology and biology are based on observations of animals in captivity. The length of the olm lifecycle also impedes research in captivity. Therefore, many questions regarding the proteus in wildlife arise.

Regarding mating, do the same partners mate in every cycle or is partner change and polygamy a frequent occurrence? If they are polygamic, does one male mate with more females during the same mating period, and *vice versa*? It is unknown whether the offspring grow up in the place they laid out, together with other siblings, or do they choose their own niche. Furthermore, the level of territoriality remains unanswered. Do adults stay in the same cave section or change their location frequently? If the location is changed, how frequent are the changes? Finally, how big are the populations in the wildlife? Are families grouped together or are they completely dispersed? Are we able to tell genealogical lineages from genotyping data?

In this master thesis, I will contribute to the knowledge of population structure and make a first parentage analysis of the olm (*P. anguinus*) population in wildlife using microsatellite data on the olms from the Postojna-Planina Cave System.

[1] *Proteus anguinus* was described in 1768, earlier than *Leptodirus hochenwartii*, but it was not recognized as cave dwelling organism. *L. hochenwartii* was recognized and described as subterranean animal in 1832.

2. AIM OF THE RESEARCH

The aim of my master's thesis is to clarify the genetic structure of the olm (*P. anguinus*) population in the Postojna-Planina Cave System. The specific aims are as follows:

1. To assess the spatial genetic structure of the olm population in the Postojna-Planina Cave System,
2. To determine conservation genetics parameter of the olm population: Hardy-Weinberg equilibrium, genetic diversity, effective number of alleles, genetic structure, and gene flow,
3. To test the possibility of determining genealogical lineages and relatedness of individuals in the population through parentage analyses.

3. MATERIALS AND METHODS

3.1. Study area

The Postojna-Planina Cave System (PPCS) represents the central area for this study. Detailed cave system structure is given in the *1. Introduction*. DNA samples of individuals analysed in this study were taken from specific parts of the PPCS: Črna Cave, Postojna Cave, and Planina Cave: Pivka Channel and Rak Channel. Samples from Črna Cave were taken from animals captured in three following cave parts: the North Tunnel, the left part of the Vilhar Tunnel, and the right part of the Vilhar Tunnel. Samples from Postojna Cave were taken from animals captured in Tartar. Samples from Planina Cave were sampled along the Pivka Channel in Planina Cave and in Rak Channel. Pivka Channel was further divided into 23 sections of length 50 metres. Additionally, olms from the Planina Channel were also captured from the downstream Pivka river pool, commonly referred to as "reserve" (slv. *rezervat*) (Figure 2).

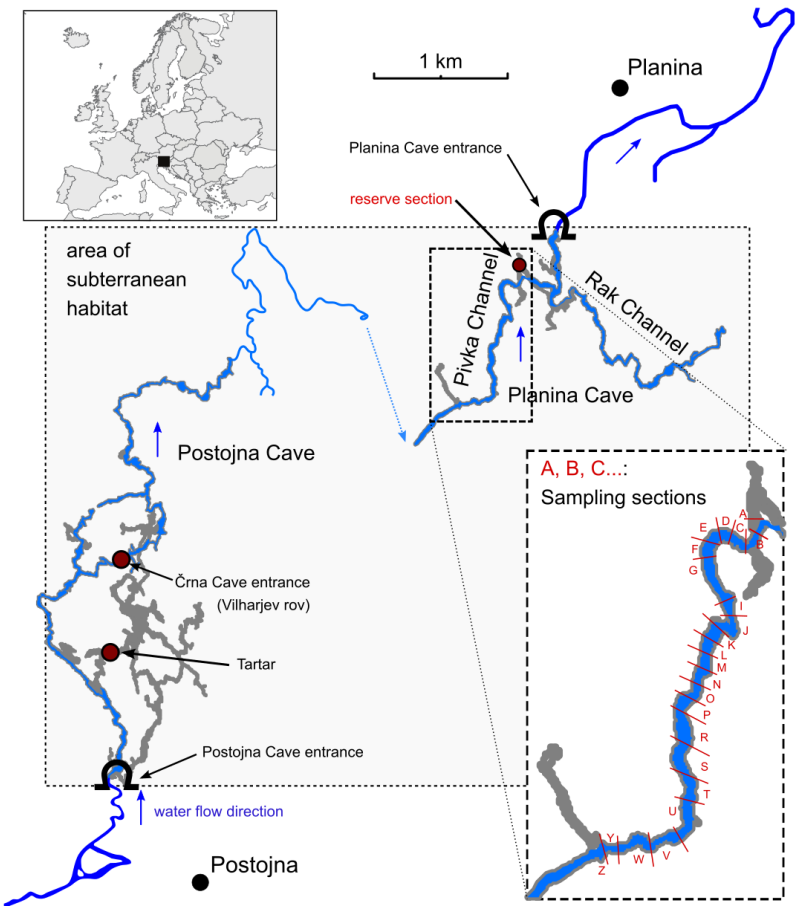

Figure 2. Map of the study area in the Postojna-Planina Cave System: Postojna Cave (consisting of Črna and Postojna Cave entrances, including Tartar) and Planina Cave (consisting of Pivka and Rak Channel). Sampling areas are marked by red dots (Črna Cave, Tartar, reserve section of Planina Cave) and by red letters (Pivka Channel). Image adjusted according to Zakšek and Trontelj (2017).

3.2. Schematic summary of the methodology

In summary, the methodology implemented in this study includes:

A. Fieldwork, apropos olm skin swab sampling in a period from 2014 to 2023;

B. Laboratory work, apropos microsatellite genotyping of swabbed individuals for 22 tetranucleotide loci;

C. Analysis of 1819 consensus genotypes, calculating population genetic parameters, designating population structure, and parentage analysis (Figure 3).

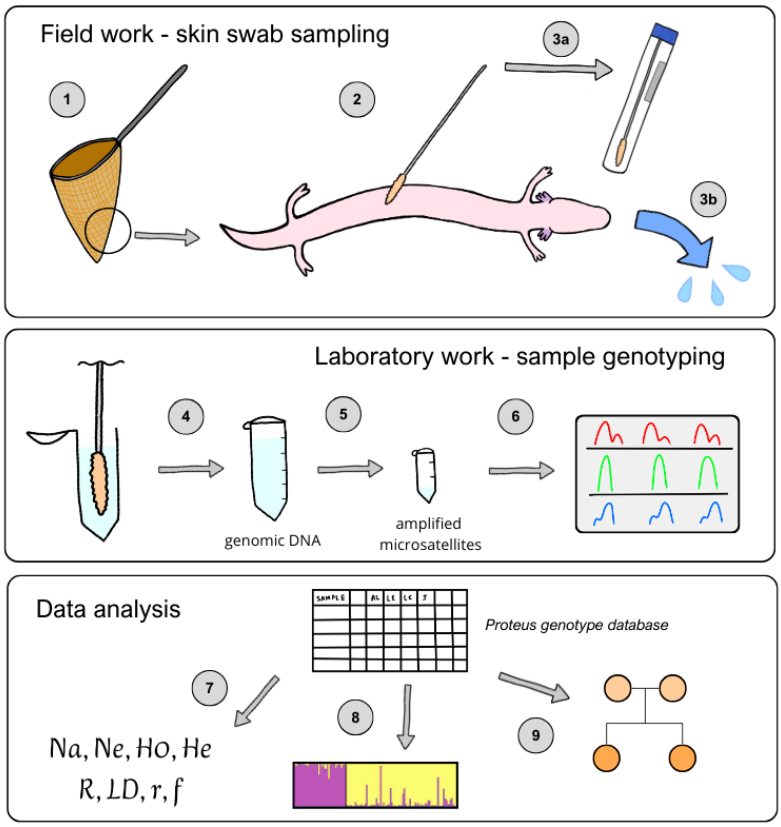

Figure 3. Schematic workflow in the study of *P. anguinus* population in Postojna-Planina Cave System.

3.3. Samples

Olms from the Postojna-Planina Cave System were caught using diving equipment and hand nets (Figure 4). Skin swab samples were taken non-invasively, from which the genomic DNA was extracted using the methodology described in Zakšek et al. (2018) and stored in the proteus DNA collection of Subterranean Biology Lab (SubBioLab) at the Department of Biology,

Biotechnical Faculty, University of Ljubljana. Most of the captured olms were also weighted and their length was measured. Extracted DNA was used for microsatellite amplification and genotyping, further explained in subdivision *3.3. Microsatellite amplification and allele length.* To test the sampling and genotyping methodology, swabs of ten individuals were taken twice on purpose. Obtained microsatellite genotypes were included in the Microsoft Access Proteus microsatellite database (further referred to as "Proteus Access Database") programmed by Tomaž Skrbinšek (Biotechnical Faculty, University of Ljubljana). Based on the loci amplification success, some of the samples were genotyped two times or more. Approximately 20% of all samples were genotyped at least twice. For the final dataset, consensus genotypes consisting out of 22 loci were calculated in the Proteus Access Database. A number of amplifications per sample ranged between 1.33 and 1.55, with a mean value of 1.43, meaning that every locus for every sample was genotyped 1.43 times on average. For my dataset, I used mainly genotypes already available in the Proteus Access Database.

Figure 4. Catching olms (*P. anguinus*) using hand nets and taking skin swabs for microsatellite genotyping.

The first sampling in PPCS started in 2014 in Črna Cave. In 2015 and 2016, two comprehensive samplings aiming to perform the mark-release-recapture method were conducted in Planina Cave (Pivka Channel) on 23 previously described sections (Zakšek and Trontelj 2017). From 2017 until 2023 some more sporadic samples were collected. In total, 1819 samples were

collected, and the majority of samples were collected in Planina Cave (Pivka Channel) (Table 1). In further parts, I will use the term "Postojna Cave" to refer to both Postojna and Črna Cave.

Table 1. Number of olm (*P. anguinus*) skin swabs sampled from 2014 until 2023 from different sites of the Postojna-Planina Cave System.

Cave entrance	year						
	2014	2015	2016	2017	2018	2021	2023
Postojna Cave	/	6	/	1	/	/	/
Črna Cave	8	33	100	/	/	/	/
Planina Cave	/	802	838	11	6	10	4
Total	1819						

3.4. Microsatellite amplification and allele length

The Proteus Access Database comprises olm genotypes consisting of 23 initial tetranucleotide microsatellite markers developed by Zakšek et al. (2018), however, one marker showed multiple and nonspecific amplifications. Therefore, it was not used in further analysis, and 22 microsatellite markers were used in my further analysis. After I checked all genotypes in the database, I ascertained which samples still had missing loci or flagged alleles (which could not be unambiguously determined in Genemapper software), hence I repeated the amplification of interrogative loci. PCR reaction mix contained 0.2 µM of each primer, 5 µL of multiplex or singleplex mastermix, 1 µL of Q solution (Type-it Microsatellite PCR kit, Qiagen), 1 µL of genomic DNA, and deionized water added to a final volume of 10 µL. Detailed information about the used multiplexes, primers, and microsatellite markers is given in Supplementary Material, Table S.1. The polymerase chain reaction (PCR) programme consisted of the following steps: initial denaturation at 95°C for 5 min, 35 cycles of denaturation at 95°C for 30 s, annealing at 60–64°C for 90 s (depending on the primers for certain loci and mix), extension at 72°C for 30 s, and terminating with final elongation at 60°C for 30 min. The PCR products were analysed using the ABI 3730 Genetic Analyzer (Applied Biosystems), using the GeneScan 500 LIZ internal size standard (Applied Biosystems). I visually checked the allele lengths of the amplified microsatellites in the Genemapper v. 6.0 program (Applied Biosystems).

3.5. Population parameters

The allelic dropout rate, false allele rate, and consensus genotypes were calculated using a Proteus Access Database. To identify genotyping errors, large allele dropouts, and the presence of null alleles in the microsatellite dataset, I used the program Micro-Checker (Van Oosterhout

et al. 2004). To determine the borderline of missing or mismatching alleles between genotypes belonging to the same individual, I run the *amUniqueProfile* algorithm included in the Allelematch package (Galpern et al. 2012). The obtained results were used to set criteria for matching analysis in the Proteus Access Database. All possible matches were manually checked and genotypes of the samples from the same individual were joined to the reference sample (animal). I calculated recapture as the ratio between the number of recaptured olms in 2016 and the number of captured in 2016 at the same section of Pivka Channel. For the analysis of body size, I compared the lengths of all captured individuals from each sampling location. For the recaptured individuals, I took the lengths of the reference sample. To calculate allele frequency, number of alleles per locus, effective number of alleles per locus, observed heterozygosity, expected heterozygosity, probability of identity, and probability of identity between siblings, I used GenAlex 6.5 package for Microsoft Excel (Peakall and Smouse 2012). To calculate allelic richness, I used FSTAT 2.9.3.2. (Goudet 2002). To estimate linkage disequilibrium, a departure from Hardy-Weinberg equilibrium, and to conduct analysis of molecular variance (AMOVA) in order to compare the proportion of genetic variation between Planina and Postojna cave, as well as Rak channel, I used Arlequin 3.5 (Excoffier & Lischer, 2010). To estimate the degree of genetic structuring and to explore population structure, I used two methodologies: the classic *F*-statistics (Wright 1931) implemented in Arlequin, followed by the Bayesian model-based clustering approach implemented in Structure 2.3.4 (Pritchard et al. 2000). I conducted analysis in Structure using admixture model without any *a priori* information, for *K* from 1 to 4, with 10 iterations. Length of the *burnin* period of 250 000, and the number of *MCMC* repeats after burnin was 1 000 000. Program is very sensitive to different sub-population sizes, which can lead to wrong conclusions if their sizes are not uniform (Puechmaille 2016; Wang 2017). Therefore, to make the numbers of samples more equalized, I ran analyses of the following datasets:

1. Separately for samples from Planina Cave and separately for samples from Postojna Cave, to distinguish for possible population structuring inside each,
2. All samples from Postojna Cave, with a corresponding number of samples from Planina Cave, samples evenly distributed along the Planina Cave.
3. 49 samples from Postojna and 46 samples from Planina Cave, both evenly distributed along Postojna Cave and Pivka Channel of Planina Cave, together with all 13 samples from Rak Channel.

Analysis (1) was conducted to check for possible population structuring inside each cave. The objective of analysis (2) was to determine possible population structuring between Postojna and Planina Cave. Analysis (3) was conducted to ascertain possible differences between the Rak Channel population and other individuals from PPCS. I run triplicate runs for analysis (2) and (3), with different samples in order to bypass possible sample bias during the sample selection. I determined the optimal number of clusters (K) using the Evanno method (Evanno et al. 2005) implemented in the CLUMPAK server (Kopelman et al. 2015).

3.6. Relatedness and inbreeding analysis

To determine the relatedness (r) and inbreeding (f) coefficients among the individuals, I used the program Coancestry 1.0.1.11. (Wang 2011). Coancestry is the most widely used software program for calculating r and f which implements seven relatedness and four inbreeding estimators. The software conducts both *a priori* simulations and analysis of empirical data. Since the ancestry of sampled olms is completely unknown, the obtained relatedness (r) cannot be compared with the real relationships. Therefore, I performed a simulation project to test for the best relatedness estimator using 200 dyads in each category: parent–offspring ($r = 0.5$), full siblings ($r = 0.5$), half siblings/avuncular/grandparent–grandchild ($r = 0.25$), double first cousins ($r = 0.25$), first cousins ($r = 0.125$), second cousins ($r = 0.03125$), and unrelated ($r = 0$). Simulations were modelled according to Taylor et al. (2015). Based on the simulation project, triadic likelihood (TrioML; Wang, 2007) and dyadic likelihood (DyadML; Wang, 2002) estimators exhibit the lowest difference from the true relatedness value. Therefore, I calculated relatedness among olms in Postojna Cave and in Planina Cave using both estimators. I used $r \geq 0.45$ cutoff to identify possible first-degree relatives: full sibling (FS), parent-offspring (PO), and $r < 0.18$ to identify non-related individuals. Relatedness cut-off was modelled based on Diez et al. (2015), from the experiment with known sibship relationships. Additionally, I identified pairs with $r \geq 0.6$ as pairs with very high relatedness values. I calculated inbreeding coefficients (f) using four available estimators: two moment estimators (Lynch & Ritland, 1999; Ritland, 1996, in the further text referred according to the authors' surnames), triadic likelihood (Wang, 2007) and dyadic likelihood (Anderson & Weir, 2007) estimators. According to Marshall et al. (2002), I labelled inbreeding coefficients of the value above 0.25 as 'high inbreeding', $0.125 \geq f < 0.25$ as 'moderate inbreeding', $0.01 \geq f < 0.125$ as 'low inbreeding', and below 0.01 as 'no inbreeding'. To determine inbreeding coefficients (F_{IS}) using F-statistics, I used the program FSTAT.

3.7. Parentage and sibship analysis

To perform parentage and sibship analysis, I used a full-pedigree likelihood methodology implemented in Colony 2.0.7.1. (Jones & Wang, 2010) and maximum likelihood methodology implemented in ML-Relate (Kalinowski et al. 2006). Colony enables the implementation of a wide variety of *a priori* information about the data, from known full-sibships to paternity and maternity (Jones & Wang, 2010). It also enables good handling of null alleles, genotyping errors, and mutations, and assigns statistical confidence for particular parent-offspring pairs (Jones et al. 2010). Finally, the program outputs a posterior probability of each sibling and parent-offspring pair. ML-Relate (Kalinowski et al. 2006), implements maximum likelihood estimates of relatedness, with the only input data being genotypes of all individuals in the dataset. As an output, ML-Relate gives the relationship (R) with the highest likelihood LnL(R) between every two individuals and specifies how lower is the log-likelihood Delta Ln(L) from other relationships. The advantage of ML-Relate is the performance of specific hypotheses tests. Program moderately handles null alleles and cannot accommodate for genotyping errors and mutations (Jones et al., 2010). Since programs assume no linked loci and no departure from HWE, I checked the given parameters in Arlequin for all analysed datasets. After multiple preliminary runs in Colony, I performed final analyses assuming male and female polygamy, dioecious diploid species, and without inbreeding, since the last parameter is chosen only when there is strong evidence of a high inbreeding level in the population. I chose a long length of run and very high likelihood precision. For every dataset, I did a triplicate analysis, with 10 runs and three different seeds chosen at random. Other parameters were set as default. Marker error rates, including allelic dropout rate and false allele rate, were calculated directly in the initial Proteus Access Database, including 1819 genotyped samples. I calculated allelic frequencies separately for Planina and Postojna Cave using GenAlex 6.5 package for Microsoft Excel. Since sex and generation in most of the olm samples aren't know, the genotype dataset couldn't be separated for offspring, males and females. To increase the informativeness of the data, I added one known full sibling pair raised in the Cave Laboratory Tular (the specimens' origin is Planina Cave). Parents of these siblings were taken from Planina Cave and the siblings were raised in captivity. All preliminary runs in Colony recognized this pair as full siblings without any *a priori* information. Furthermore, when I chose this pair as an already known sibling pair in preliminary runs, other possible full sibling pairs were more consistent among consecutive runs. As a mismatch threshold for both father-offspring and mother-offspring dyads, I chose three loci, to account for possible errors and mutations. There was no excluded

paternity, maternity, nor paternal or maternal sibships. As an input, the ML-Relate program demands only the genotypes of all analysed samples, without any *a priori* data on relationships among individuals. To run the maximum likelihood analysis, I chose a 95% confidence interval and 100 000 seed.

As we don't know much about "true sibship" in the population, we select a strict criteria and treat only relationships that were consistent in both analyses (Colony and ML-Relate) as reliable. In case ML-Relate results differ from Colony results, I run the specific hypothesis tests implemented in ML-Relate. Furthermore, all identified relationships were checked with biological data we have: In case the result from ML-Relate did not agree with the biology of the samples (e.g. subadults smaller than 21 cm recognized as parent and offspring), more focus was given to the results obtained by Colony.

I conducted exploratory analyses of parentage for three smaller datasets:

1. Samples taken from the Rak Channel,
2. Samples taken from the L section in Pivka Channel,
3. Samples from the entrance parts of Planina Cave: sections called reserve, A, B, and C.

Samples from Rak Channel were chosen due to their smaller sample size, separate location in the cave, indicated differences in population structure, and due to higher relatedness compared to individuals from other parts of the Planina Cave. In section L, three juvenile individuals (body length of 12.5 – 13.5. cm) were caught close to each other, which prompted the parentage and sibship analysis of this section. Due to the calculation barriers of available infrastructure, it was not possible to analyse datasets having more than 100 individuals. Therefore, samples from the entrance parts of the Planina Cave were chosen in order to cover a bigger area compared to the first two analysis, but simultaneously using a dataset with less than 100 samples.

4. RESULTS

4.1. Initial dataset and detection of individuals

During the sampling from 2014 until 2023, 1819 skin swabs of *P. anguinus* from the Postojna-Planina Cave System were collected (Figure 5). Genotyping failed for 16 samples, hence the final dataset had 1803 consensus proteus genotypes (22 loci). Loci PA32, PB21, and PC32 were the only loci with an amplification success rate below 90%. Error rates included false allele rate and allelic dropout rate. Most of the loci had a false allele rate below 0.01%, whereas locus PA32 manifested the highest false allele rate (0.13%). Loci PA22 and PA32 were the only loci with allelic dropout rate above 2% (2.07% and 2.05% respectively). No genotyping errors or large allelic dropouts were detected when analysing set of 1803 genotypes. The highest number of alleles was observed for loci PA22 and PB03 (12 alleles). The mean number of different alleles per locus is 7, while the mean effective number of alleles per locus is 2.08.

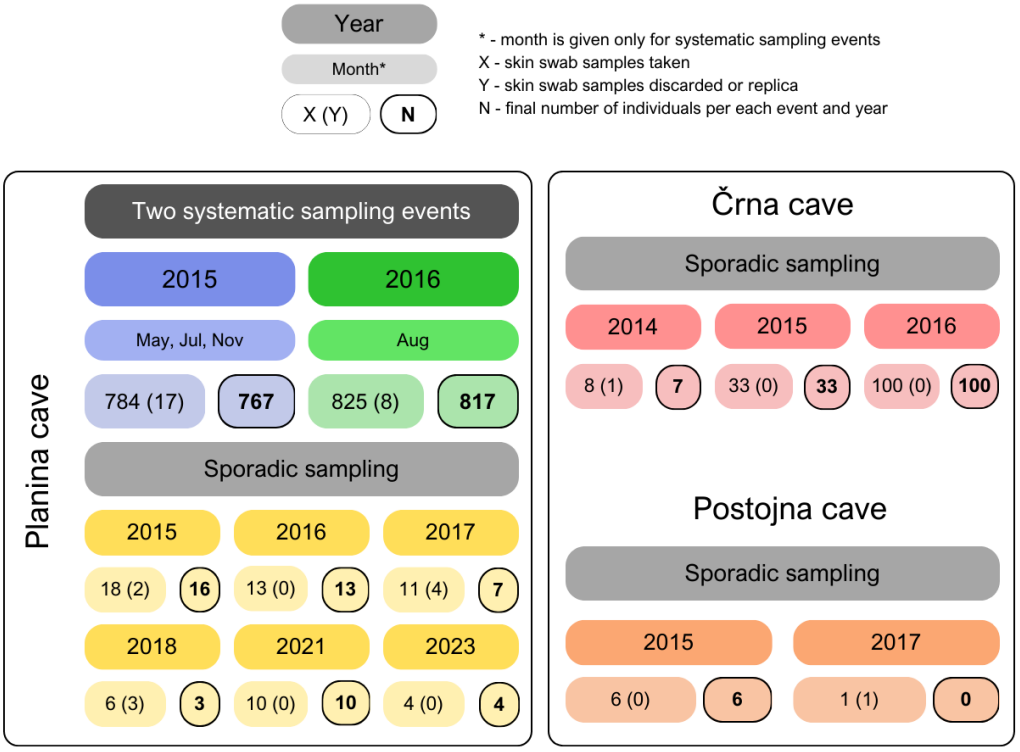

Figure 5. Overview of sampling events from different sites of the Postojna-Planina Cave System. Events are partitioned into two systematic sampling events during 2015 and 2016 for mark-release-recapture study, and other sporadic sampling events from 2014 until 2013. Planina Cave and Postojna Cave (including Postojna and Črna Cave entrances) are visually partitioned on the scheme. The number of olm (*P. anguinus*) skin swab samples taken, discarded, and the final number of individuals per each event and year are given. All 135 matching samples are included.

Low probability that two individuals within a population share the same genotype ($P_{ID} = 1.8 \times 10^{-11}$) and low probability that two siblings share the same genotype ($P_{IDsib} = 1.2 \times 10^{-5}$) enable individual identification using the given microsatellite marker set. To identify the samples that have the same genotype (and should belong to the same individual), I run matching in the Proteus Access Database. In order to account for genotyping and mistype error, I used the criteria where: maximum two missing alleles (allelic dropout) or not more than one mismatch was allowed to identify samples which probably originate from the same individual.

The matching resulted in 156 matching dyads from 146 reference samples. Out of them, 20 dyads represented replica samples, either taken to test methodology (n = 14) or sampled two days in a row from neighbouring locations (n = 6). After the removal of the mentioned samples, matching results included 135 dyads from 133 reference samples.

After identification of the same individuals, I removed duplicate genotypes from the dataset for further analyses (population genetic parameters, analysis of population structure, parentage analysis). The final dataset resulted in 1647 individuals.

4.2. Population genetic parameters

Population genetic parameters were calculated separately for the samples from Postojna Cave (Črna and Postojna Cave) and Planina Cave (Pivka and Rak Channel). Observed (H_o = 0.090 – 0.692 for Planina, H_o = 0.125 - 0.685 Postojna) and expected (H_e = 0.100 – 0.712 for Planina, H_e = 0.120 – 0.701 for Postojna) heterozygosity were relatively low for microsatellite markers (Table 2). Average H_o and H_e is similar for both caves (H_o =0.450 ± 0.166, H_e = 0.459 ± 0.172 in Postojna Cave; H_o =0.448 ± 0.180; H_e = 0.459 ± 0.184 in Planina Cave). Four loci from Postojna Cave population and nine loci from Planina Cave population are not in Hardy-Weinberg equilibrium ($p < 0.05$). Among individuals captured in Rak Channel, two loci were monomorphic (all individuals have the same allele), all loci were in HWE, and all loci had less alleles compared to Postojna Cave and Pivka Channel. Observed (H_o = 0.154 – 0.769, mean H_o = 0.486 ± 0.205) and expected (H_e = 0. 222 – 0.760, mean H_e = 0.493 ± 0.178) heterozygosity for samples captured in Rak Channel was similar to the values obtained for Postojna and Planina Cave (Table 3). Allelic frequency for both caves, a complete list of private alleles, and allelic frequencies of Rak Channel samples are given in Supplementary material (Tables S.2., S.3. and S.4).

Table 2. Population genetic parameters of olm (*P. anguinus*) populations in two caves (Postojna and Planina) for 22 microsatellite loci. Abbreviations: *n* – number of analysed individuals, N_a – number of detected alleles for each locus, N_e – effective number of alleles for each locus, H_o – observed heterozygosity, H_e – expected heterozygosity, *R* – allelic richness, based on minimum 143 individuals, * - indicates departure from Hardy-Weinberg equilibrium ($p < 0.05$).

Locus	Postojna cave						Planina cave					
	n	***N_a***	***N_e***	***H_o***	***H_e***	***R***	***n***	***N_a***	***N_e***	***H_o***	***H_e***	***R***
PA01	144	6	1.5	0.382	0.362*	5.419	1502	7	1.2	0.173	0.181*	5.698
PA02	144	4	2.2	0.514	0.535	4.257	1500	5	2.6	0.592	0.615	4.236
PA12	144	4	1.4	0.340	0.324	4.166	1503	6	1.3	0.246	0.251	4.205
PA21	144	3	2.2	0.576	0.544	3.728	1503	4	2.2	0.542	0.543	3.694
PA22	144	9	2.5	0.653	0.599	7.859	1502	12	1.7	0.392	0.407	8.560
PA31	144	4	2.1	0.528	0.536	3.372	1503	6	1.5	0.315	0.320	3.419
PA32	144	6	1.8	0.417	0.435	6.698	1449	8	3.0	0.616	0.671*	6.812
PB01	144	8	2.7	0.542	0.627*	9.161	1502	11	2.8	0.632	0.641*	9.255
PB02	144	6	2.2	0.563	0.555	4.780	1502	6	1.5	0.344	0.350*	5.074
PB03	144	7	1.5	0.299	0.309	7.951	1498	12	2.5	0.611	0.600	7.922
PB12	143	7	3.3	0.685	0.701	7.280	1497	9	3.3	0.680	0.695*	7.258
PB21	144	6	2.8	0.597	0.649	7.659	1486	10	3.5	0.692	0.712	7.567
PB22	144	4	1.2	0.125	0.120	4.452	1496	5	1.6	0.382	0.388*	4.485
PC01	144	4	2.4	0.576	0.593	4.751	1503	8	2.2	0.554	0.553	4.698
PC02	144	5	1.2	0.160	0.164	4.486	1499	6	1.7	0.430	0.429	4.615
PC03	144	3	1.6	0.438	0.382*	3.000	1503	3	1.1	0.120	0.123	3.000
PC11	144	6	1.2	0.167	0.180	5.093	1492	6	1.1	0.090	0.100*	5.358
PC12	143	4	2.9	0.594	0.662	4.315	1501	6	2.5	0.580	0.601*	4.353
PC21	144	6	2.3	0.569	0.563	4.643	1503	6	2.1	0.513	0.524	5.237
PC22	144	4	2.0	0.507	0.489	3.454	1491	4	2.5	0.584	0.595	3.510
PC31	144	4	2.1	0.424	0.512*	3.791	1503	5	1.7	0.367	0.398*	3.890
PC32	144	3	1.4	0.236	0.252	4.183	1495	5	1.6	0.403	0.387	4.154

Table 3. Population genetic parameters of olm (*P. anguinus*) population in Rak Channel (Planina Cave) for 22 microsatellite loci. Abbreviations: *n* – number of analysed individuals, N_a – number of detected alleles for each locus, N_e – effective number of alleles for each locus, H_o – observed heterozygosity, H_e – expected heterozygosity, *R* – allelic richness, based on minimum 12 individuals.

Locus	Rak Channel					
	n	***Na***	***Ne***	***Ho***	***He***	***R***
PA01	12	4	1.48	0.333	0.424	4.000
PA02	13	3	2.86	0.692	0.532	3.000
PA12	13	4	1.59	0.231	0.222	3.769
PA21	13	4	2.15	0.462	0.711	3.923
PA22	/	1	1	/	/	1.000
PA31	13	2	1.75	0.462	0.443	2.000
PA32	13	3	2.70	0.769	0.665	3.000
PB01	13	5	2.91	0.615	0.729	4.920
PB02	13	3	1.48	0.154	0.151	2.846
PB03	13	5	2.86	0.615	0.668	4.917
PB12	13	5	2.82	0.769	0.760	4.923
PB21	13	4	2.72	0.692	0.618	3.920
PB22	13	4	1.27	0.385	0.458	3.994
PC01	13	3	2.43	0.308	0.385	2.923
PC02	13	3	1.62	0.231	0.280	2.923
PC03	/	1	1	/	/	1.000
PC11	13	3	1.23	0.538	0.615	3.000
PC12	13	3	2.64	0.692	0.600	3.000
PC21	13	2	1.91	0.538	0.409	2.000
PC22	13	3	2.14	0.692	0.520	2.923
PC31	13	3	1.35	0.385	0.397	3.000
PC32	13	2	1.50	0.154	0.271	2.000

Linkage disequilibrium (LD) was calculated for the two datasets: (i) samples from Postojna and Planina Cave (including Rak Channel), (ii) samples distributed in Postojna Cave, Planina Cave (without Rak Channel), and Rak Channel ($p < 0.01$) (Supplementary material, Table S.5.). For dataset (i), 14 and 42 pairs of loci were found to be linked for Postojna and Planina cave, respectively. For dataset (ii), 14, 32, and 2 pairs of loci were found to be linked for Postojna cave, Planina cave, and Rak channel, respectively. In Planina cave, the highest number of linked loci was connected to the locus PB21.

4.3. Overview of recaptures

Systematic sampling conducted during May, July, and November 2015 and August 2016 in Planina Cave resulted in 767 and 817 animals sampled each year. In each cave section, a similar number of individuals was caught in 2015 and 2016 (Figure 6). Most of the olms were captured in F, H, and Z sections, while the least number were captured in the A and O sections. Overall, the lowest number of olms were captured in the cave sections closer to the entrance of Planina Cave (A, B, C, D, E).

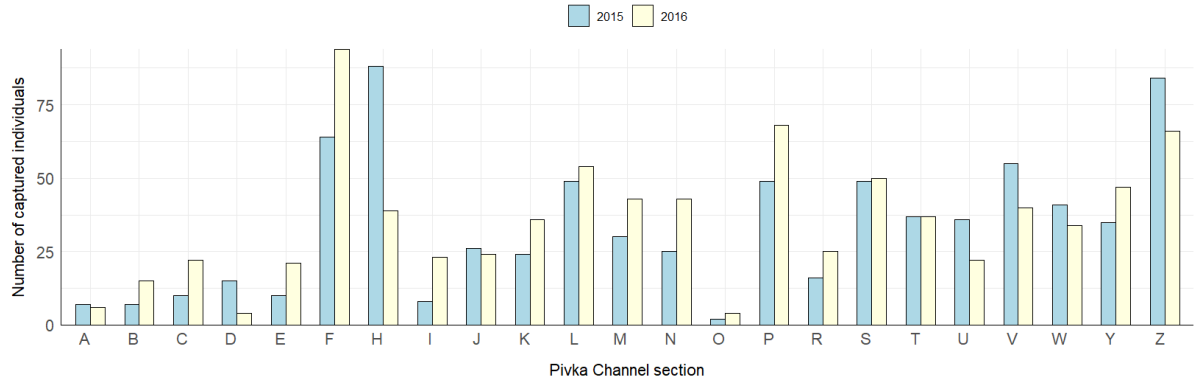

Figure 6. Number of captured olms (*P. anguinus*) on each section of Pivka Channnel in Planina Cave during systematic sampling in 2015 and 2016.

Altogether, 115 individuals were recaptured in 2016, which gives the recapture ratio of 14.1% (Figure 7). Recapture ratio was the highest for section B, however, the low number of initially captured individuals at this section contributes to the size of the obtained number. The highest number of recaptures was in the F section. There were no recaptured individuals on the sections A, C, K and O. Recapture rate in other sections do not deviate much from the average recapture ratio.

93 individuals (80.9%) of recaptured olms, were recaptured on the same cave section indicating high site fidelity of recaptured individuals. The rest of recaptured individuals were recaptured either on the closest neighbouring section (11, 9.6%, e.g., L and M sections), second neighbouring section (8, 7.0%, e.g., D and F sites), third neighbouring section (2, 1.7%, e.g., C and F sites) or fourth neighbouring section (1, 0.9%, e.g., N and S sites) (Figure 8).

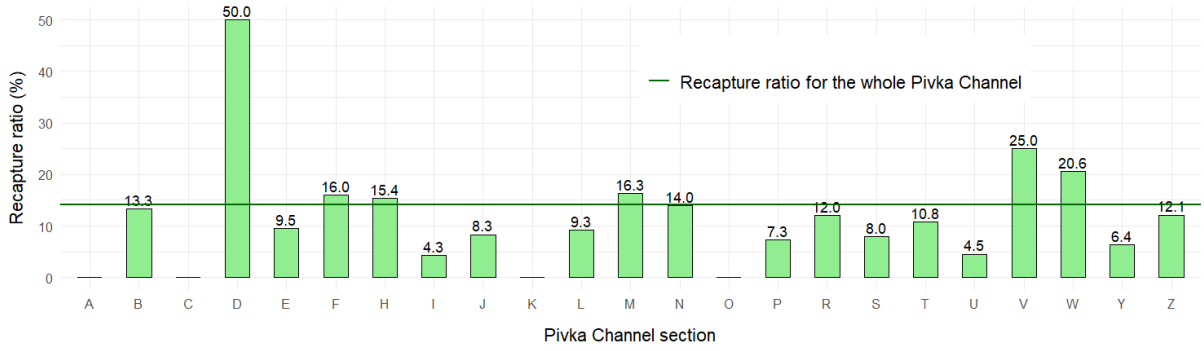

Figure 7. Recapture ratio of olms (*P. anguinus*) for each Pivka Channel (Planina Cave) section, with ratio values above the bar. Dark green line shows the recapture ratio for the whole Pivka Channel (14.1%).

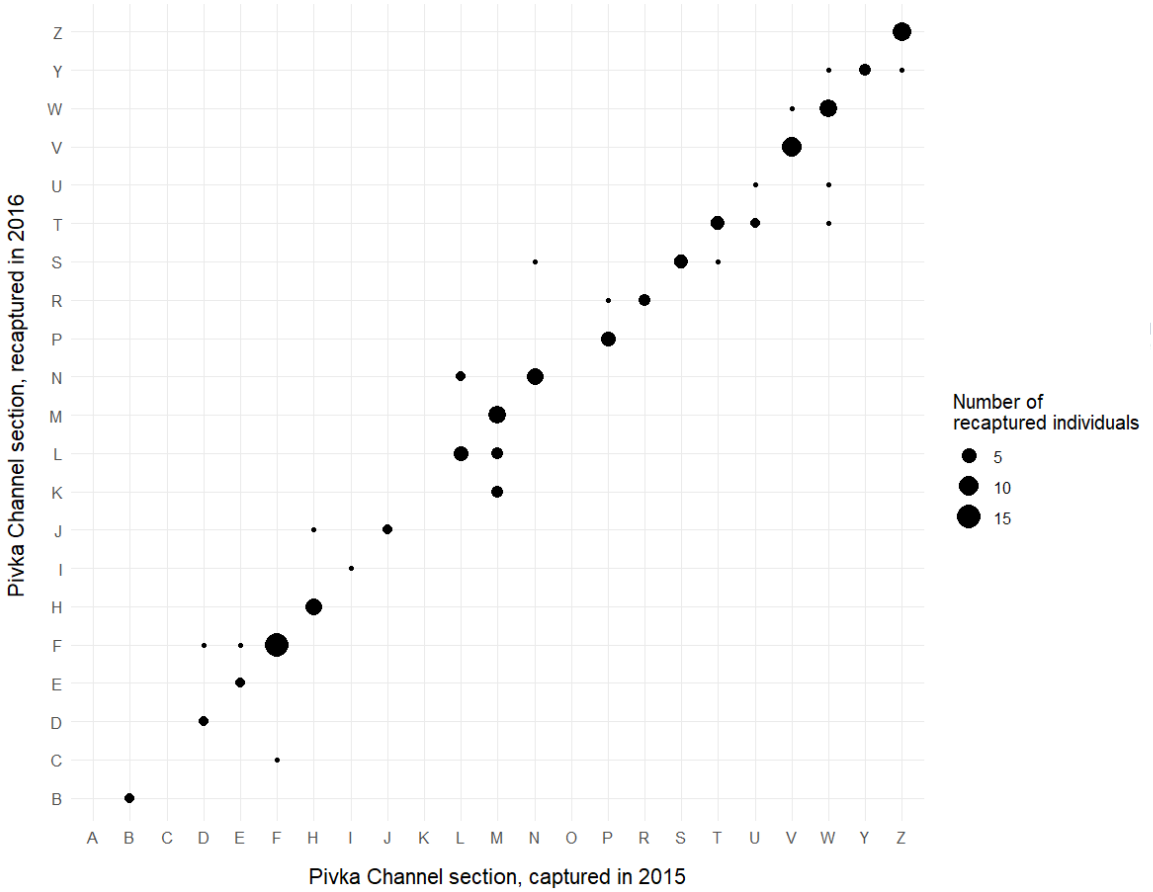

Figure 8. Overview of olm (*P. anguinus*) recapture locations along Pivka Channel (Planina Cave). The x-axis represents the location where olms were captured in 2015, and the y-axis represents the corresponding recapturing location in 2016.

4.4. Recaptured olms between 2014 - 2023

Besides the olms that were recaptured during the extensive and systematic sampling in 2015 and 2016, some samples were more randomly collected before and after this period (between 2014 and 2017). Here, I show some interesting recaptures (with sample voucher code numbers in brackets):

- Olm recaptured after 4.5 months (PA890), firstly on section I, afterward in section W positioned on the upstream section of Planina Channel;
- Olm recaptured after 1 year 7 months (PC306), for the first time captured on the W section of Planina Channel, afterward in the "reserve" section downstream side of Planina Cave;
- Olm recaptured after 2 years 1 month (PC570), firstly on B, afterward in A section of Planina Channel;
- Olm recaptured three times (PB809), firstly after 1 year 1 month on the same location (I), secondly 2 years 2 months after the first capture on the second neighbouring location (G);
- Olm recaptured after 2 years 3 months (PA484), firstly on H, afterward in the G section of Planina Channel;
- Olm recaptured after 8 years 1 month (PC997), both times in A section of the Planina channel. It was 12.7 cm long when captured , and 5.8 cm longer when recaptured;
- Two olms have been recaptured after 1 year, both times in the right part of the Vilhar Tunnel (Črna Cave), being the only two individuals recaptured in Črna or Postojna Cave.

4.5. Morphological characteristics of sampled olms

Most of the captured olms were longer than 20 centimetres (Figure 9), hence representing adults, mature and in a stage when able to reproduce. The mean length of olms in Postojna Cave is: 24.51 cm (Q_1 = 23 cm, M = 25 cm, Q_3 = 26.63 cm), in Planina Cave 23.30 cm (Q_1 = 22.5 cm, M = 23.5 cm, Q_3 = 24.5 cm), and in Rak Channel 17.58 cm (Q_1 = 14.5 cm, M = 18.5 cm, Q_3 = 19.5 cm).

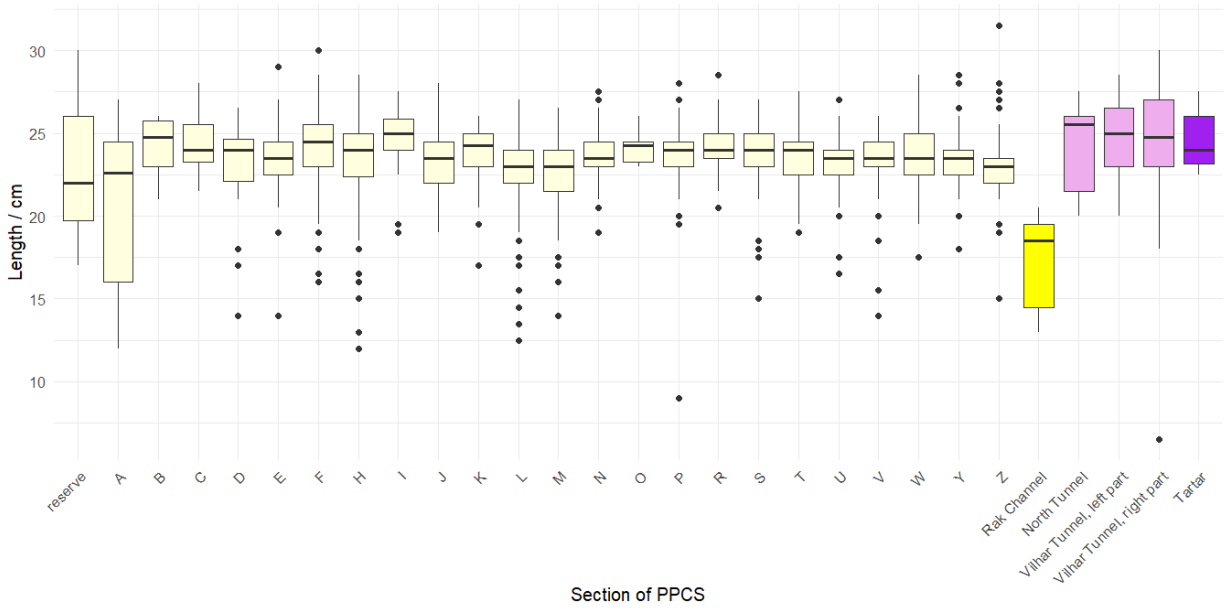

Figure 9. Length of the olms (*P. anguinus*) captured in different parts of Postojna-Planina Cave System between 2014 and 2023. Yellow colours correspond to Planina Cave (n = 1503), and violet colours correspond to Postojna Cave (n = 144). Legend: light yellow – Planina Channel, with reserve, yellow – Rak Channel, violet – Črna Cave (North Tunnel, Vilhar Tunnel), dark violet – Postojna Cave (Tartar).

4.6. Population structure

AMOVA analysis across 22 microsatellites shows genetic variability between Planina (n = 1503) and Postojna Cave (n = 144). The percentage of variation among populations and within populations are 7.89% and 92.11 %, respectively. Classical F-statistics also indicates genetic structure (F_{ST} = 0.07889, $p < 0.05$).

With samples from Rak Channel being separated from samples from Pivka Channel of Planina Cave, AMOVA and classical F-statistics show higher genetic variability between three populations (Table 4, 5).

Table 4. Percentage of variation among three putative olm (*P. anguinus*) populations in PPCS. Number of samples: Postojna Cave – 144, Pivka Channel (Planina Cave) – 1490, Rak Channel (Planina Cave) – 13.

	Postojna Cave	Pivka Channel	Planina Channel
Postojna Cave	/		
Pivka Channel	7.99 %	/	
Rak Channel	12.04 %	13.89 %	/

Table 5. F_{ST} - values among three putative olm (*P. anguinus*) populations in PPCS. Number of samples: Postojna Cave – 144, Pivka Channel (Planina Cave) – 1490, Rak Channel (Planina Cave) – 13.

	Postojna Cave	Pivka Channel	Planina Channel
Postojna Cave	/		
Pivka Channel	0.7986	/	
Rak Channel	0.12041	0.13893	/

Structure analyses of all genotyped individuals from Planina Cave (Figure 10.A) and Postojna Cave (Figure 10.B) using the Bayesian model-based clustering approach indicate a lack of genetic structure inside the Planina and Postojna caves, respectively. However, Rak Channel samples show slightly different pattern compared to the rest of the Planina Cave samples.

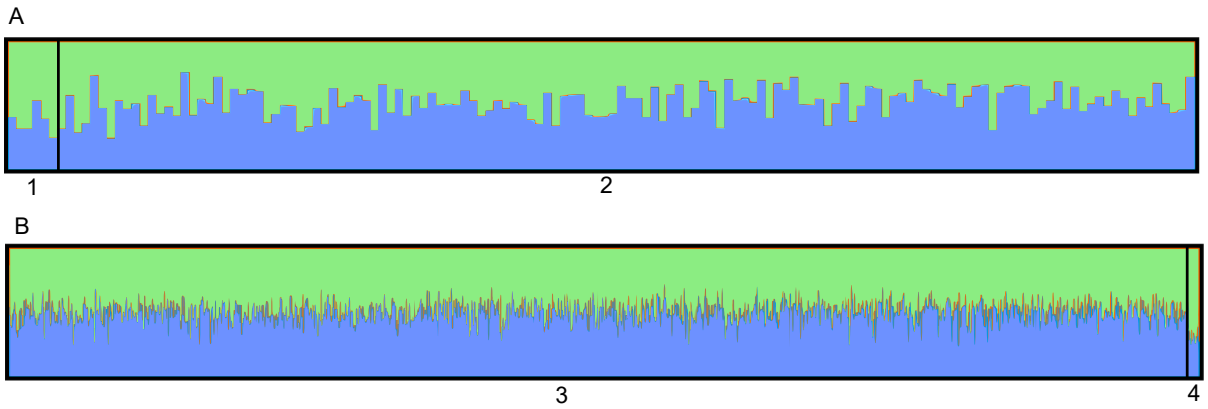

Figure 10. Structure genetic cluster analysis of olms (*P. anguinus*) from Postojna (A) and Planina (B) Caves into groups (K = 2) using 22 microsatellite loci without prior information on the location of individuals (n = 1647). Each individual is represented by one column divided according to its probability of membership of "one population": blue population and green population. A - Postojna Cave (6 samples from Postojna (1) and 138 samples from Črna Cave (2)). B - Planina Cave (1490 samples from Pivka Channel (3) and 13 samples from Rak Channel (4)).

Structure genetic clustering indicates a genetic structure between Postojna and Planina Cave (Figure 11). Most of the individuals from Postojna Cave (sections "1" and "2" in Figure 11.A) have the highest probability of membership to the "purple" cluster, whereas most of the individuals from Planina Cave (section "3") have the highest probability of membership to the "yellow" cluster. Individuals from the Rak Channel of Planina Cave (section "4") show different patterns compared to the rest of the Planina Cave, as it is indicated by the partition into three groups (Figure 11.B). However, these Structure plots indicate some genetic structuring between Postojna and Planina Cave, and some samples show more similarities to the samples from another cave. I made triplicate runs with different samples from Planina cave, taking an equal number of individuals from each part of Planina and Rak channel (Supplementary Material, Figure S.1.). The best partition for all three analysed sample groups by the Evanno method is K = 2.

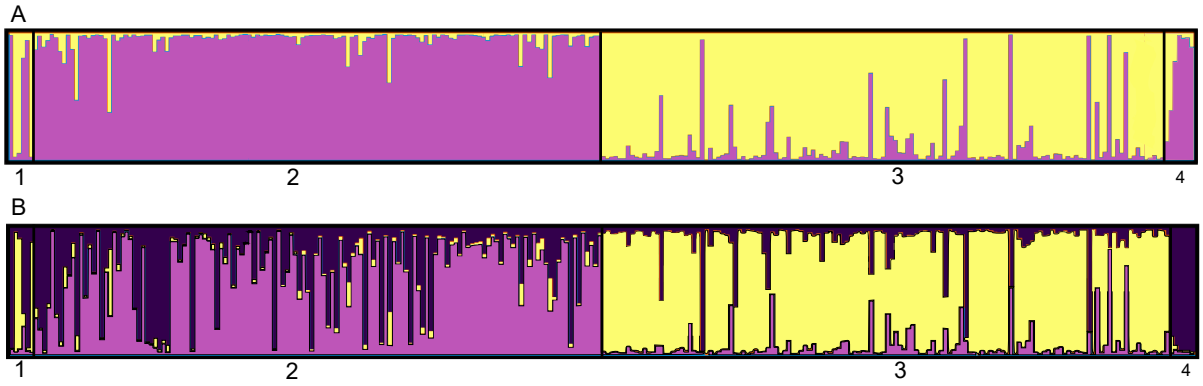

Figure 11. Structure genetic cluster analysis of olms (*P. anguinus*) from Postojna (1, 2) and Planina (3, 4) caves into groups: $K = 2$ (A) and $K = 3$ (B) using 22 microsatellite loci without prior information on the location of individuals ($n = 288$). Each individual is represented by one column divided according to its probability of membership of "one population": violet population and yellow population (A), and additionally indigo population (B). First 144 samples represent samples from Postojna Cave (6 samples from Postojna (1) and 138 samples from Črna Cave (2)). The second 144 samples represent samples from Planina Cave, six samples randomly taken from each cave section. The order of Planina channel sections from left to right is A – Z (3), ending with Rak Channel (4). Triplicate runs were performed, with more detail in Supplementary material, Figure S.1.

Bayesian model-based clustering approach from Structure indicates that the Rak channel population slightly differs from all other samples in the PPCS (Figure 12). In order to overcome the program's bias for unequal sample size, I lowered the number of samples from the Postojna Cave and Planina Channel and took into account all genotypes from Rak Channel. Analogously, I made triplicate run with different samples from the Postojna and Planina Cave (Supplementary Material, Figure S.2.). The best partition for all three analysed sample groups by the Evanno method is $K = 2$.

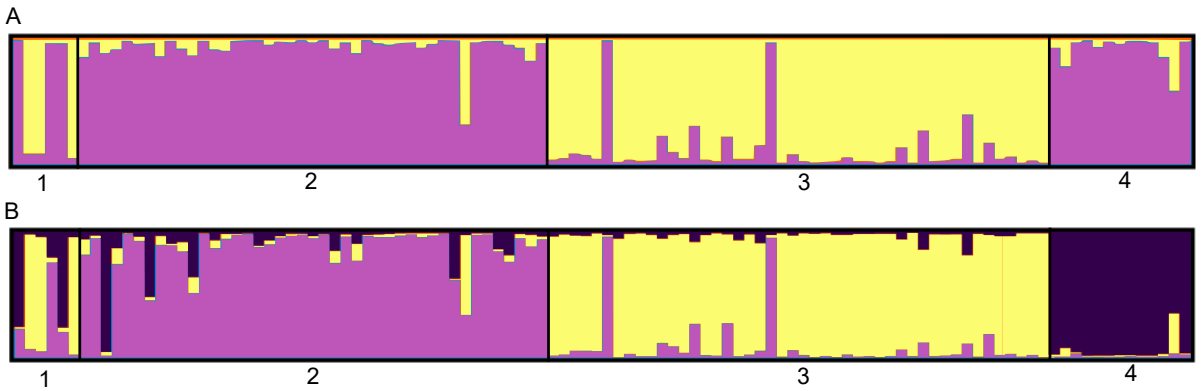

Figure 12. Structure genetic cluster analysis of olms (*P. anguinus*) from Postojna (1, 2) and Planina (3, 4) caves into groups: $K = 2$ (A) and $K = 3$ (B) using 22 microsatellite loci without prior information on the location of individuals ($n = 108$). Each individual is represented by one column divided according to its probability of membership of "one population": violet population and yellow population (A), and additionally indigo population

(B). The first 49 samples represent samples from Postojna Cave (6 samples from Postojna (1) and 43 samples from Črna Cave (2)). The second 59 samples represent samples from Planina Cave (46 samples from Pivka (3) and 13 samples from Rak Channel (4)). The order of Planina channel sections from left to right is A – Z (3), ending with Rak Channel (4). Triplicate runs were performed, with more detail in Supplementary material, Figure S.2.

4.7. Putative relatedness

We studied relatedness within each location of PPCS (Postojna, Pivka Channel and Rak Channel) and putative relatedness between all three parts. Triadic and dyadic likelihood estimators outperformed other relatedness estimators in simulation analysis, with a 79% similarity between real relatedness values and calculated relatedness values in simulated genotypes. Therefore, relatedness values among sampled individuals were calculated using both estimators (Table 6). In both simulations and real data estimates, the dyadic likelihood estimator was less conservative compared to the triadic likelihood estimator, producing more dyads with higher relatedness values. Individuals from Rak Channel (13 ind.) exhibit higher relatedness estimates among each other compared to the relatedness estimates among other individuals from Planina Cave (Figure 13, showing only representative sample of Pivka Channel using more conservative, e.g. triadic likelihood, approach; Supplementary Material Figure S.3). Among the samples from the Postojna Cave, no group differed from the others according to the estimates of relatedness.

Table 6. Estimates of relatedness value (*r*) among 1503 olms (*P. anguinus*) from Planina Cave (Pivka and Rak Channel), 144 individuals from Postojna Cave and 13 individuals from Rak Channel using two methods: triadic likelihood (TrioML) and dyadic likelihood (DyadML) estimators. Meaning of relatedness abbreviations: VH – very high related individuals, with a relatedness value above 0.6; H – highly related individuals, with a relatedness value between 0.45 and 0.6; R – related individuals, with a relatedness value above 0.18 and below 0.45. Each relatedness group is connected to the number of pairs having given relationship (in brackets) and the number of individuals included in those pairs. The average value of relatedness estimates is given in the last column.

Planina Cave				
r (relatedness)	VH	H	R	average
	$r \geq 0.6$	$0.6 > r \geq 0.45$	$0.45 > r \geq 0.18$	
TrioML	162 (92)	1443 (10101)	1503 (105259)	0.0824
DyadML	393 (276)	1471 (12220)	1503 (143977)	0.0977
Postojna Cave				
r (relatedness)	VH	H	R	average
	$r \geq 0.6$	$0.6 > r \geq 0.45$	$0.45 > r \geq 0.18$	
TrioML	14 (7)	101 (151)	144 (1045)	0.0616
DyadML	27 (14)	105 (169)	144 (1313)	0.0713
Rak Channel				
r (relatedness)	VH	H	R	average
	$r \geq 0.6$	$0.6 > r \geq 0.45$	$0.45 > r \geq 0.18$	
TrioML	3 (2)	9 (8)	13 (57)	0.3115
DyadML	4 (4)	9 (8)	13 (54)	0.3241

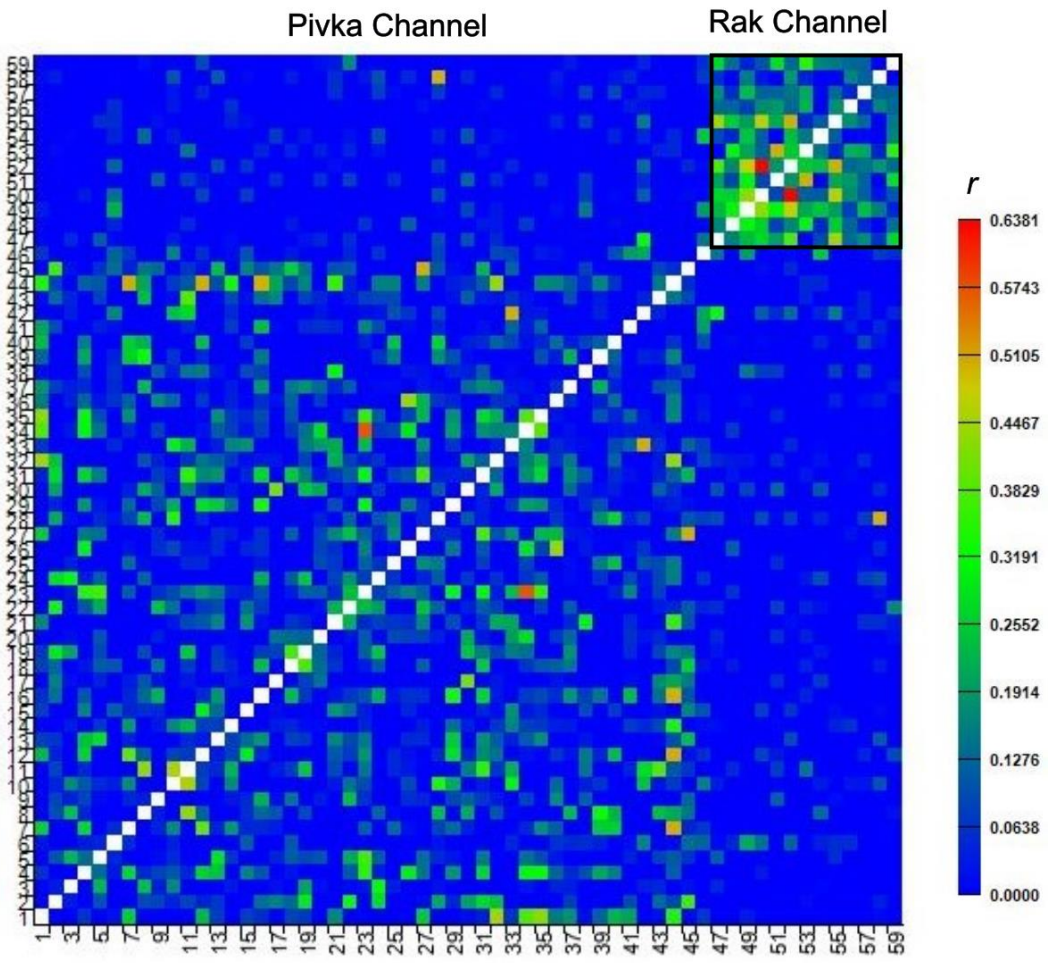

Figure 13. Estimates of relatedness (*r*) among olms (*P. anguinus*) sampled from Planina Cave (Pivka and Rak Channel) using triadic likelihood estimator. The analysis included 46 individuals evenly distributed across the Pivka Channel and 13 individuals from Rak Channel. X and Y axis represent individuals, while each square on the plot represents the estimated relatedness value (e.g. the colour matching the legend) for given dyad, corresponding to individuals from X and Y axis. The data above the diagonal line ($y = x$, in white) correlates with the data below the line. Relatedness among samples from Rak Channel is squared in the upper right corner of the picture. The relatedness plot for the whole Planina Cave is in Supplementary Material, Figure S.3.

4.8. Inbreeding

Inbreeding coefficients, estimated in Postojna and Planina Cave populations, showed inbreeding in olm populations (Figure 14). Two different types of moment estimators (named by the authors: (i) Ritland, (ii) Lynch & Ritland) produced similar estimations (Figure 14.A,B), as well as two different likelihood estimators (named by the methods: (i) dyadic, DyadML, (ii) triadic, TrioML) (Figure 14.C,D). However, inbreeding coefficients calculated using two main methodologies differ from each other. Calculated estimation of inbreeding for individuals from both caves (Postojna Cave and Planina Cave, including Rak Channel) were higher when obtained by likelihood estimators compared to moment estimators. Moment estimators predicted that more than 50% of sampled individuals are not inbred, while likelihood estimators assume this number is around 30%. Based on likelihood estimators, most of the olms exhibit low inbreeding levels. Obtained F_{IS} values for Postojna Cave, Planina Cave (including Pivka and Rak Channel), and Rak Channel were: 0.02, 0.023, and 0.015, respectively.

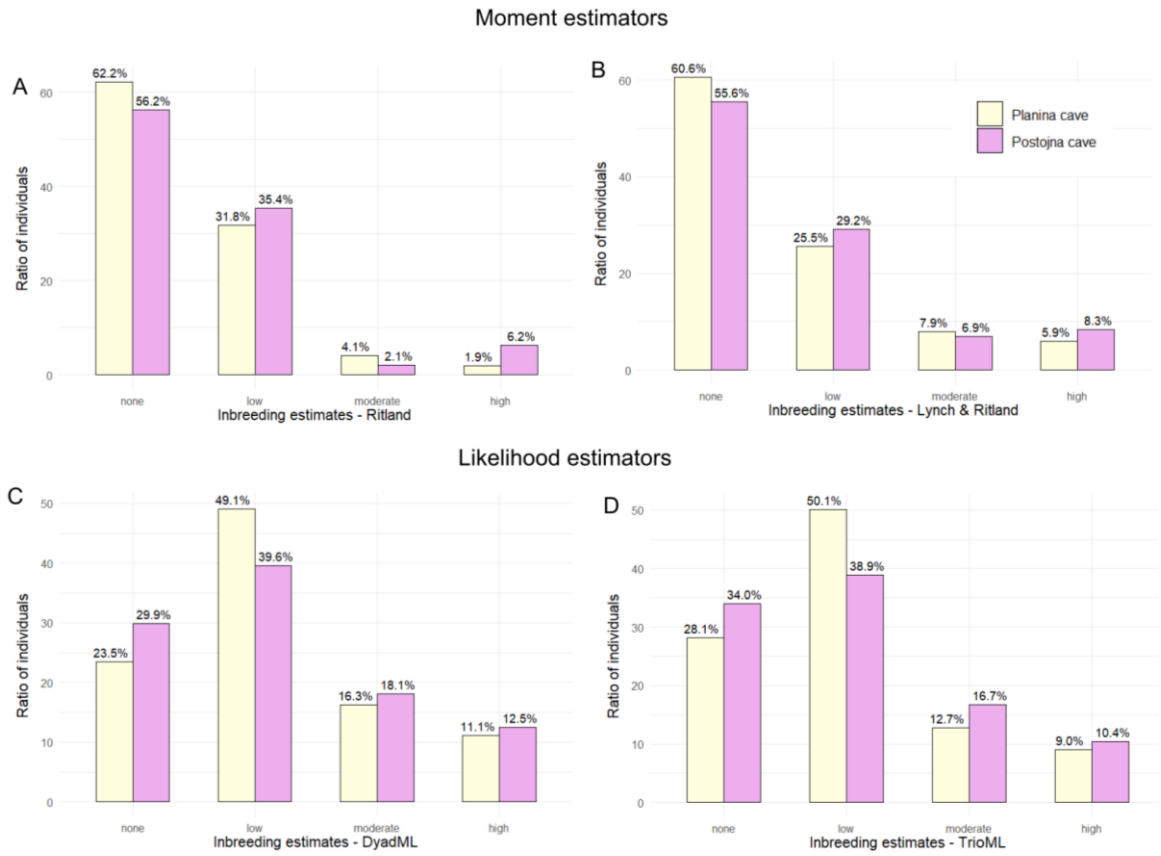

Figure 14. Estimates of inbreeding in the olm (*P. anguinus*) populations from Planina and Postojna Cave using two moment estimators (A – Ritland, B – Lynch & Ritland) and two likelihood estimators (C – DyadML, D – TrioML). Inbreeding coefficient (f) between 0 and 0.01 is labelled as no-inbreeding ("none"), $0.01 \geq f < 0.125$ as "low", $0.125 \geq f < 0.25$ as "moderate" and $f \geq 0.25$ as high inbreeding ("high"). The total number of analysed individuals from Planina Cave is 1503, while from Postojna Cave is 144.

4.9. Parentage and sibship

For detailed parentage analysis using both full-pedigree likelihood (implemented in Colony software) and maximum likelihood estimator (implemented in ML-Relate software) I chose three different datasets in Planina Cave: Rak Channel, L section of Pivka Channel, and entrance sections of Planina Cave (reserve, A, B, C). Samples from Rak Channel were selected due to their geographic isolation from the Pivka Channel, their distinct population structure predicted by theBayesian clustering approach, AMOVA and classical *F*-statistics, and higher relatedness compared to individuals from other sections in the Pivka Channel (Planina Cave). Three juvenile individuals captured in the L section were staying close to each other before sampling, prompting the sibship analysis of this section. Computing limitations did not enable the analysis of datasets with more than 100 individuals. Consequently, samples from the entrance regions of the Planina Cave were selected to cover a larger area than the Rak Channel and L section, with less than 100 samples.

All analyses were conducted using the additional full sibling (FS) pair known from captivity in Cave Laboratory Tular, and their parents were taken in Planina Cave. This FS pair was included to control and calibrate the precision of both programs and a variety of analyses with different parameters.

None of the parentage analyses for the sample set in the Rak Channel, L section, and entrance parts of Planina Cave conducted by the full-pedigree likelihood method indicated parent-offspring (PO) relationships between the sampled olms. However, analyses revealed potential full-sibling (FS) and half-sibling (HS) pairs. Analyses of the same datasets using the maximum likelihood approach resulted in all three types of relationships, including PO. Number of HS and FS pairs was higher compared to the number obtained by the full-likelihood method for L and entrance sections of Planina Cave. Potential relationships which were not concordant between different methods, or which were biologically questionable were additionally tested through specific hypothesis tests (Table S.13.).

4.9.1. Sibship among individuals from Rak channel

Sibship among olms in the Rak Channel (Planina Cave) was reconstructed on genotypes of 13 individuals. The length of individuals was smaller compared to the rest of the PPCS, with an average length of 17.58 cm (Q_1 = 14.25; M = 18.5; Q_3 = 19.5). The full pedigree likelihood method recognized four potential FS pairs (Supplementary Material, Table S.6.). The maximum likelihood method did not recognize any FS pair among individuals from the Rak channel. On the other hand, three out of these four FS pairs that were recognized by full-likelihood method, maximum likelihood method recognized as PO pairs, with specific hypothesis tests not excluding these relationships (Supplementary Material, Table S.7.). According to the body length of individuals, the FS relationship between these individuals seems to be more likely than PO as all individuals are smaller than 20.5 cm and therefore probably juveniles or young specimens. The fourth FS pair suggested by the full-likelihood method was assigned to the HS pair when using maximum likelihood method. Since the results from the full pedigree likelihood method couldn't be compared with the results from the maximum likelihood method, the full pedigree likelihood probability borderline of 0.5 (e.g., the method assigned this pair the given relationship in more than 50% runs) was set to decide which half sibling (HS) relationships are more reliable.

In addition to these four FS pairs, the full pedigree likelihood method suggested HS relationships in another six individuals within 13 samples from Rak Channel (Figure 15Figure 10). However, the full likelihood method suggested a known full-sibling pair from Cave Laboratory Tular, included in the analysis as a full-sibling control pair, as closely related to individuals from Rak Channel. Since this full-sibling pair was raised in captivity, the suggested HS relationship with the population in Rak Channel is less likely.

Results of suggested relationships between individuals in Rak Channel are variable between different methods, and the only result that is consistent is that approximately half of the captured olms in Rak Channel are related to each other (FS, HS, PO).

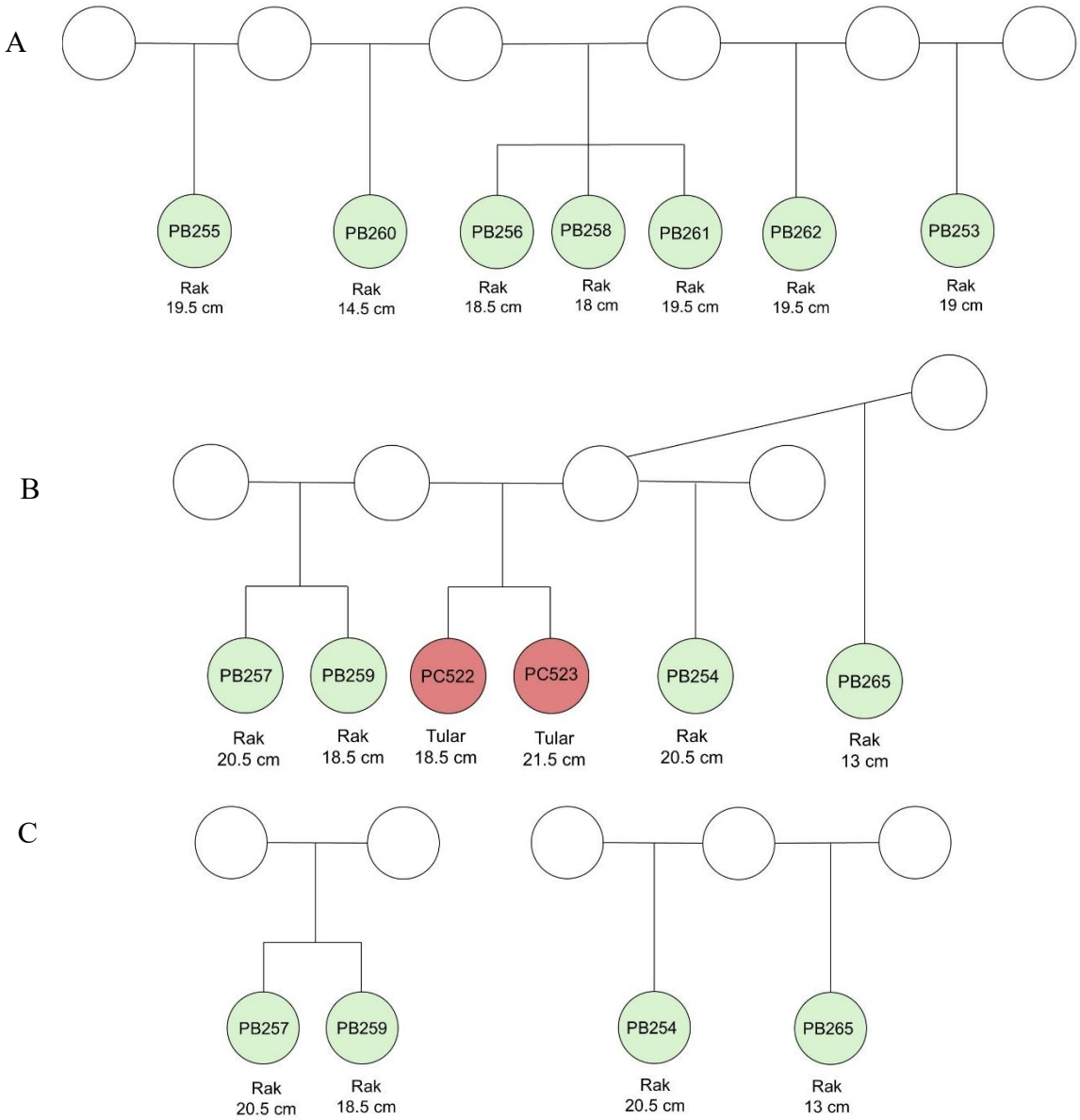

Figure 15. Genealogical trees of olms (*P. anguinus*) in Rak Channel (Planina Cave) constructed using the full-pedigree likelihood method, including (A & B) and excluding (A & C) individuals from Cave Laboratory Tular in the figure. According to the full-pedigree likelihood method, individuals from the genealogical tree (A) are not related to Tular individuals. Coloured circles represent offsprings with genotypes of analysed samples and corresponding sample voucher codes, with the colour matching the location of capture: green – Rak Channel, red – Cave Laboratory Tular (captivity). The body length of each individual is listed beneath the circles. White circles represent the parents of individuals.

4.9.2. Sibship among individuals from entrance parts of Planina Cave

Sibship among olms in the entrance parts of Planina Cave was reconstructed on 84 genotypes of individuals captured (or recaptured) in sections: reserve, A, B, or C of Pivka Channel. The full pedigree likelihood method suggested six FS pairs with very high probability (Supplementary material, Table S.6.). All and the same FS were recognized by the maximum likelihood method too. Five FS pairs were captured in the reserve section of Planina Cave, while only one FS pair was suggested and captured in the C section. Moreover, both methods suggested five individuals who are in half-sibling relation to the suggested full-sibling pairs (Figure 16). Suggested HS were captured in all sections of entrance part of Planina Cave. Four individuals with assigned sibships were recaptured. All of them were recaptured on the same section where they were captured for the first time, with the recapture after 1 month up to 2 years.

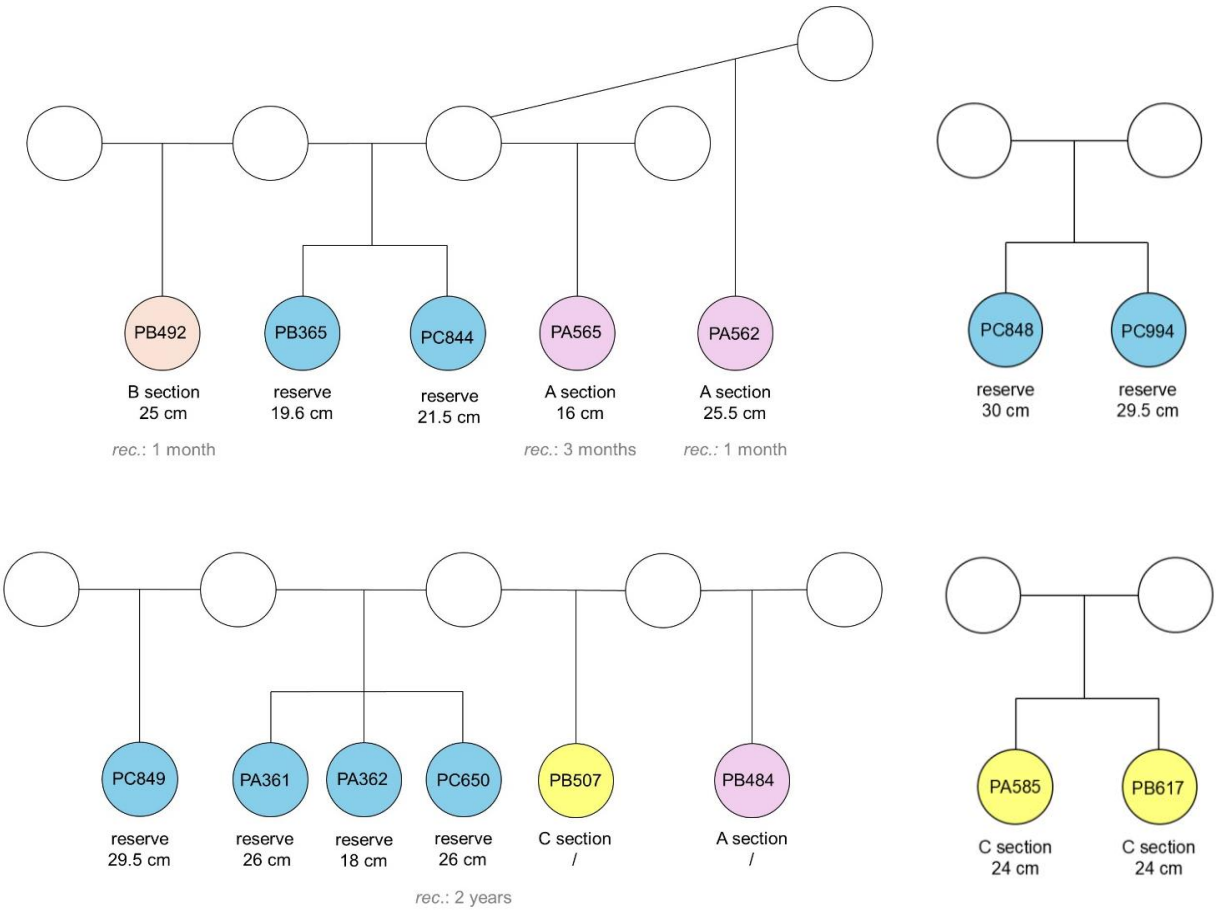

Figure 16. Genealogical trees of olms (*P. anguinus*) from entrance sections (reserve, A, B, C) of Pivka Channel (Planina Cave) suggested by full-pedigree and maximum likelihood methods. Coloured circles represent offsprings with known genotypes and corresponding sample voucher code, with the colour matching the location of capture: blue – reserve section, pink – A section, orange – B section, yellow – C section. The body length of each individual

is listed beneath the circles. If the individual was recaptured, recapture time (*rec.*) is written beneath the body length. White circles represent parents of individuals.

4.9.3. Sibship among individuals from L section in Pivka Channel

Sibship among olms in the L section of Pivka Channel (Planina Cave) was reconstructed on 99 genotypes of individuals captured (or recaptured) on L section. The full pedigree likelihood method suggested six full sibling (FS) pairs. Out of these six FS pairs, maximum likelihood method suggested three FS and three parent-offspring (PO) pairs (Supplementary Material, Table S.5.). For one suggested pair (PB178 and PC000), specific hypothesis test could not assign any relationship (FS, PO) as a better fit (Supplementary Material, Table S.7.). Looking at this pedigree from the broader perspective, individuals PB178 and PC000 have the additional shared FS, individual PB983. In this case, the relationship obtained by the full-pedigree method fits the biological data better, hence the pair was concluded to be FS. Another FS pair not suggested as FS by maximum likelihood (ML) method are two out three juveniles captured together (body length between 12.5 - 13.5 cm) and wild guess on the field if they might be related originally initiated the analysis of L section. The maximum likelihood method suggested these three juveniles as parent and offspring, and the specific hypothesis test did not exclude this relationship as possible. However, regarding their body length, the PO relationship of these individuals is highly unlikely. The third FS pair which was not consistently suggested by both methods was not accepted as plausible via hypothesis testing. Altogether, only two individuals who are half-siblings of the recognized full-siblings were supported by both methodologies (Figure 17).

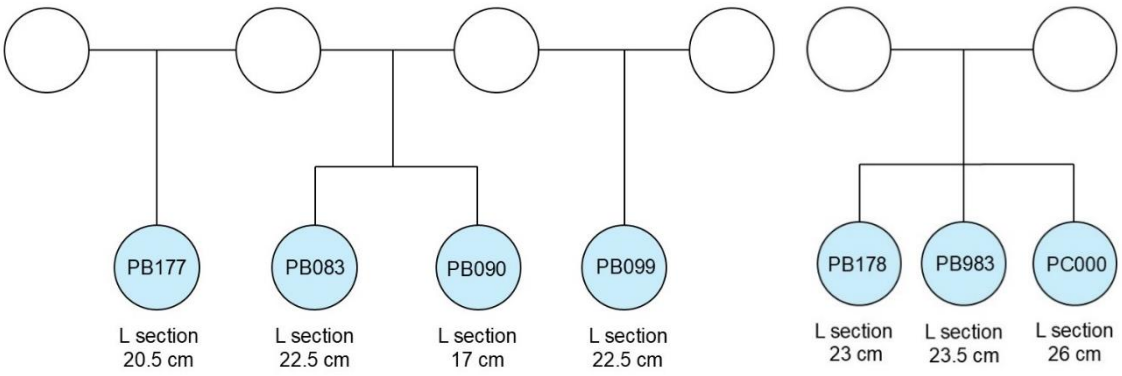

Figure 17. Genealogical trees of olms (*P. anguinus*) from the L section of Planina Cave constructed using both full-pedigree and maximum likelihood methodologies. Blue circles represent offsprings with known genotypes and corresponding sample voucher codes. The length of each individual is listed beneath the circles. White circles represent the parents of individuals. Since the sex of each individual is unknown, all individuals are represented as circles.

5. DISCUSSION

5.1. Population parameters and population structuring

Population genetic parameters of olms within PPCS using microsatellites were already studied in 2018 where 201 individuals from Postojna and Planina Cave were included (Zakšek et al. 2018). The study of this master thesis includes a much higher number of individuals sampled along Pivka Channel (Planina Cave) and Postojna Cave, and in addition samples from another water flow - Rak Channel (Planina Cave), therefore complementing the sample size and sites included in the previous study.

Population genetic parameters of the larger dataset showed higher linkage disequilibrium (LD) compared to the smaller sample set in the previous study and more loci that did not confirm Hardy-Weinberg equilibrium expectations in Postojna and Planina Cave. By removing the samples from Rak Channel out of the Planina Cave population, the number of linked loci decreased by 23.8%.

The first analysis of population structure inside PPCS showed weak genetic structuring and nearly panmictic genetic pattern (Zakšek et al., 2018). Here, I showed that the genetic structure between olms in Postojna Cave and Planina Cave is stronger than detected by a smaller dataset and that olms from Rak Channel (Planina Cave) are slightly different from the other sites in PPCS. In addition, in this study variance detected by AMOVA was higher (7.89% compared to 0.24% in the previous study) and F_{ST} values were higher (0.07886 compared to 0.0024 in previous study). Furthermore, genetic differences in samples from the two caves were shown by Structure plots and the Evanno partition method.

Sampling bias that occurred with the analysis of only smaller area of PPCS (Zakšek et al. 2018), together with the high sensitivity of the Structure software (Pritchard et al. 2000; Puechmaille 2016) to the number of samples in each population very likely influenced the previous analysis and result. Skin swabs from the last study were taken only from the downstream areas of Planina Cave (reserve, A, B, C, D, E, F, and H part of Planina Channel; data from Proteus Access Database) and only a few samples were taken from Postojna Cave (including Črna Cave). Therefore, samples analysed by Zakšek et al. (2018) were not collected along the entire length of Postojna and Črna Caves, nor along the entire length of Planina Cave. These led to an underestimated observed population structure between the two caves, although the underlying genetic structure is much stronger. Furthermore, deviations from HWE and linkage between

loci might be caused by inbreeding, population stratification or non-random mating. However, these effects still need to be clarified.

The samples from Rak Channel show genetic clustering and can be distinguished from Postojna and Planina Cave samples by AMOVA, and classical *F* – statistics, and are distinguishable by a Bayesian model based structuring approach. Relatedness and parentage analyses of the samples from Rak Channel indicate high relatedness and family relationships between individuals in this small sample, which might have an impact on the results. Anderson & Dunhan (2008) found that the Structure software is very sensitive to data containing family members, recognizing them together as a separate population. Consequently, the suggestion of partition of Rak Channel samples into separate sub-populations could be the result of family lineages, leading to greater similarity between the analysed individuals. Furthermore, the number of analysed individuals ($n = 13$) is small and could have an impact on allele frequencies. Therefore, these results should be considered with some caution and for a clearer picture of the genetic parameters of the olms from Rak channel, it is necessary to increase the sample size. However, the olms in Rak Channel are not so numerous and easily accessible which makes sampling there much more challenging.

5.2. Recaptures and migrations

The similar number of olms captured in each section of the Pivka Channel (Planina Cave) in 2015 and 2016 (767 and 817, respectively) indicates a similar number of individuals present in the open and accessible part of the Pivka Channel. The recapture of six individuals the next day after the first sampling in the next section indicates that the sampling method used did not stress the olms so much that they would have changed their original position significantly or found an inaccessible hiding place after release.

The recapture ratio of 14,1% along the Pivka Channel indicates that a large number of individuals were not in an assessable location during the sampling period. Caves are geographically and geologically very complex ecosystems, large parts of which are inaccessible to humans (Gunn 2004). Some of the narrow passages have not even been explored due to physical obstacles. Regarding the low recapture olms, the olm population studied therefore consisted of animals that live more or less permanently in wide passages and animals that might change their locations. More than 90% of the recaptured olms were found on the same or a neighbouring section of the Pivka Channel, suggesting that at least some of the animals in the population exhibit high site fidelity. The site fidelity of olms was studied in Vruljak I Cave in

Bosnia and Herzegovina by Balázs et al. (2020), where the animals were individually tagged with a unique black colour pattern in a 270 m long section starting from the cave entrance. After the first capture, the tagged individuals were followed for 28 months during four diving expeditions. Out of 19 tagged individuals, 13 (e.g., 2/3) were recaptured. Furthermore, out of the seven individuals tagged in the pilot study (Balázs et al. 2015), five were recaptured during the eight-year study period. A very high recapture rate in Vruljak I Cave compared to the relatively low recapture rate in Planina Cave could indicate that the population inhabiting wider passages in Vruljak I Cave is smaller than the one in Planina Cave. Furthermore, the study area in Pivka Channel of Planina Cave was larger than in Vruljak I Cave, being more than 2 kilometres along 23 sections compared to 270 m. A larger study area increases the possibility that some animals will move out of reach or hide during the capture. In Vruljak I Cave, no recaptured individual was observed more than 80 m from the capture and tagging area, which also suggests high site fidelity (Balázs et al. 2020).

Within all animals recaptured between 2014 – 2023, only two individuals were recaptured at a more distant location, indicating longer migration. One migrated more than one kilometre upstream of the Pivka River within four months, which means that this adult olm had to swim actively more than one kilometre upstream. Another adult, which was an animal in good condition according to its weight (36 grams), was recaptured approximately two kilometres downstream of the Pivka River after 18 months. One possibility is that the river current carried the animal downstream, but no other individual was observed in a similar situation. Another possibility is that this olm was actively swimming in downstream, similar to the first one recorded upstream. However, due to the low recapture rate in this study, it is not possible to determine whether large-scale migrations are rare in general. For comparison, Balázs et al. (2020) reported that a longer time period between captures did not claim longer distance moved.

The most interesting is a discovery of a young olm, which was first caught in 2015, with the body length of 12.7 cm. This individual was recaptured after more than eight years in the same section of Planina cave (A) and was almost six centimetres larger than before. This finding represents one of the longest observations of olm growth in nature. Balázs et al. (2020) reported a similar finding – they recaptured the same individual on the same location 7 years after the first capture, inside Vruljak I Cave. No size or weight information was given for this individual. Based on studies in captivity (Guillaume, 2000), it is known that olms are regularly tied to familiar places. This finding was first confirmed for olms living in nature in Bosnia and

Herzegovina (Balázs et al., 2020) and has now also been confirmed for the population within the Postojna-Planina Cave System.

Only two olms were recaptured inside the Črna Cave (belonging to the Postojna Cave population). Altogether, smaller number of samples were collected from this cave and Postojna Cave than from Planina Cave. A very low number of recaptures in Črna Cave indicates that a larger number of individuals are beyond the reach of researchers and the method used.

Since caves represent extreme and very specific habitats, the observed high site fidelity should be placed in the context of subterranean animals. The Italian Cave Salamander (*Speleomantes italicus*) showed a high site fidelity towards the cave which they use as an underground shelter. *S. italicus* exhibits outdoor activity, and regardless of other caves in a near vicinity (Tuscan Apennines), they always return to the same shelter cave (Lunghi and Bruni 2018; Lunghi et al. 2022). A mark-release-recapture study of the Ozark cavefish (*Amblyopsis rosae*) resulted in recapturing 50% of the tagged fish in one cave in Arkansas. Most of the recaptured cavefish (65%) moved less than 100 metres from the initial capture location. A few moved long distances, with the maximum distance being around one kilometre (Brown and Johnson 2001). What causes the preference towards high site fidelity among subterranean organisms is not clear.

For an additional insight into the olm population in PPCS regarding site fidelity, it would be interesting to repeat sampling and analyse population in next years with additional systematic sampling along Pivka Channel. This would enable the comparison of recapture rate and site fidelity after a decade.

5.3. Relatedness and inbreeding among olm populations in PPCS

The relatedness of olms in PPCS is high. Almost all individuals (more than 90%) from Planina Cave and a large proportion (more than 70%) of individuals from Postojna Cave included in this study probably have a first-degree relative, e. g., parent, sibling or offspring, among the analysed samples. The relatedness of individuals from Planina Cave shows a higher relatedness between individuals in Rak Channel compared to the other parts of Pivka Channel, although the sample from Rak channel could be highly biassed due to the small sample size and the fact that the samples were taken only once on a very limited area. In Postojna and Planina Cave, pairs of individuals with relatedness above 0.6 were found. Since the estimate of relatedness value among non-inbred individuals should not exceed 0.5 for the first-degree relatives, a higher value could indicate the presence of inbreeding in the population. The estimates of inbreeding

coefficients are consistent with this indication, suggesting inbreeding at least some level of inbreeding. However, different methods for estimating inbreeding do not concur with the inbreeding level in the populations. Moment estimates show far less inbreeding level in both cave populations compared to the likelihood methods. Taylor (2015) analysed the performance of estimators implemented in Coancestry software for populations having both high and low allelic diversity. At low marker diversity, unrelated and highly related individuals will have only slight disparity in genetic similarity. By analysing relatedness and inbreeding on a large sample size, estimates are expected to be more robust. Consequently, these estimators may be applied to derive mean values of *r* and *f* across the population, however they cannot be used for dyads and individuals (Taylor, 2015). As both expected and observed heterozygosity in proteus microsatellite markers are low, the relatedness and inbreeding values obtained should not be used for comparisons among individuals, but as an overview of the populations.

5.4. A trial to assess family relationships

To assign family relationships between olms in PPCS, two methods were used: full-pedigree likelihood method and maximum likelihood method. Full-pedigree likelihood method infers parentage and sibships concurrently, evaluating the likelihood over the entire pedigree configuration (Jones and Wang 2010). On the other hand, maximum likelihood method calculates the likelihood of relationships for each pair of individuals (dyads) separately (Kalinowski et al. 2006).

Apart from the theoretical and statistical background behind the two methods for parentage and sibship analysis, the input data differ among the two programs with the implemented methods. Colony (full-pedigree likelihood estimator) can implement *a priori* data with known relationship, which enable the calibration of analysis with known full-sibling pair from Cave Laboratory Tular. However, no known and genotyped parent-offspring pair was available for this study, so this couldn't be implemented as *a priori* data in the program, possibly influencing no obtained PO relationship among analysed individuals. Moreover, predictions of genotyping and mutation errors are also included in Colony and not in ML-Relate. ML-Relate (maximum likelihood estimator) allows specific hypotheses tests to be performed, thereby analysing the dyad of interest deeply and clearly. Furthermore, the microsatellite data used showed different results for different datasets. The presence of multiple loci that could be linked and are not in HWE represents a possible drawback of the proteus dataset and emphasis the need to carefully review all the results.

5.4.1. Sibship analysis of olms from Rak Channel

As a geographically separated group, concurrently having the highest level of relatedness among the individuals, Rak Channel samples served as one of the exploratory groups. Nevertheless, the results are inconsistent and not congruent and therefore challenging to interpret. There are three concerns to be discussed: (i) full-pedigree likelihood method and maximum likelihood method did not suggest any common full-sibling pair, with maximum likelihood method not suggesting any full-sibling pair; (ii) full-pedigree likelihood method suggested samples from Rak Channel to be closely related to two siblings from captivity Cave Laboratory Tular; (iii) maximum likelihood method suggest relationship that are highly unlikely regarding body size of animals.

To start with (i), relatedness values obtained for individuals from Rak Channel implies the higher level of relatedness, e.g. more individuals are having full- and half-siblings compared to other sections of Pivka Channel (Planina Cave). Results obtained by Colony concur with this implication, with full pedigree likelihood method determining full-sibling pairs. Nevertheless, the only pair recognized by maximum likelihood method implemented in ML-Relate was the control pair from Cave Laboratory Tular, therefore no full-sibling pair among Rak channel individuals was identified. The big difference between two methods questions the ability to distinguish parentage and sibship relationships in other datasets. Secondly (ii), the ancestry of control full-sibling pair from Cave Laboratory Tular - they were born in captivity, having their parents also in captivity. Parent olms were taken from Planina Cave (specific location unknown), making half-sibling relationship with the samples in Rak Channel highly unlikely. Thirdly (iii), results obtained by maximum likelihood method, implemented in ML-Relate are not very likely when body size of individuals was checked. Although individuals with the length between 18 and 20 cm represent adults, they haven't yet started to mate, hence it is unlikely that they represent each other's parents and offsprings. Since sampled individuals in Rak channel differ on many levels (geographically, different hydrographical water flow, $F_{ST,}$ Bayesian clustering approach) compared to other parts of Pivka Channel, one can discuss whether these individuals morphologically and reproductively differ from the other individuals in Pivka Channel and have already started to reproduce despite the smaller stature. However, individuals from Rak Channel were captured as a group, rather than as solitary individuals. In nature, grouping behaviour was only found among juvenile olms, while adult individuals, especially males, are living solitary life (Balázs et al., 2020). According to observation in this study, maximum likelihood results which predict parent-offspring relationship, are less likely.

Finally, this result stresses out the importance of good dataset for valid parentage and sibship analysis. As an essential desire for credible results, one strives for a larger number of samples and highly informative markers. When these criteria aren't completely met, it is important to look the wider picture and question the biological relevance of the results obtained.

5.4.2. New insights into family relationships among olms from the entrance part and L section of Planina cave

Methods used for parentage and sibship assignment suggested more consistent and congruent for analysed individuals in Pivka Channel compared to the individuals in Rak Channel. On sections with about hundred individuals, sibship analysis reveal two or three full siblings at the most. Given that females in captivity usually lay a much larger number of eggs, from 20 to 60 (Juberthie et al., 1996), there are several possible explanations for suggested number of siblings: (i) the number of eggs laid in wildlife is much smaller, (ii) the survival and hatching of offspring in wildlife is much lower, (iii) siblings disperse throughout the cave during development. All recognized FS pairs came from the same section, but it is important to state that maximum four sections were analysed simultaneously (out of 24 in Pivka Channel of Planina Cave) and that low recapture indicates much bigger population than sampled. Furthermore, the dataset included low number of juveniles. In the nature, the probable solution is a combination of all three assumptions.

Apart from denoting FS pairs, methodologies suggested HS from both parents' sides. This implies that both females and males can mate with more than one partner during the lifecycle. Some of the FS pairs significantly differ in the size (e.g. PA361 and PA362 from L section), and therefore represent individuals of different ages, indicating that the same partners can mate in more mating seasons. All suggested FS were captured on the same section, however individuals from HS pairs were not necessarily captured on the same section, implying that families can be spread across the area of at least 200 m.

Three juvenile individuals from the L section were captured together, but no FS nor HS pairs were confirmed by both FL and ML methodologies. Nevertheless, sibship analysis implies that two of them might be related, however specific relationship couldn't be determined. The result implies that offsprings stick together after hatching, even if they are not from the same parents. They might be grouped by hatching location. Afterwards, offsprings might disperse to different parts of the pit and occupy a new niche. To investigate the last assumption, the sibship among all sampled individuals should be analysed together.

5.4.3. The future challenges in olm parentage analysis

This study represents the first attempt to check for relationships among olms. Sex identification in olms is difficult due to the absence of sexual dimorphism and homomorphic sex chromosomes and there were already several attempts to develop a non-destructive and reliable approach for sex identification (Gredar et al. 2019). Method to determine gender of olm is still unknown, representing a big challenge for population studies (Flanagan and Jones 2019), since the lack of gender information is a big problem in researching genealogical relationships. More accurate parentage analyses would be possible with the possibility to distinguish among two groups: potential parents and potential offsprings, as well as with the set of data with known relationships. While distinguishing among potential parents and offsprings is impossible for the olm populations from the nature, getting the set of data with known relationship feasible. Keeping and raising individuals in captivity represents a great opportunity to obtain samples of individuals with known parentage and sibship. In this trial, technological limitations of available infrastructure permitted only the usage of smaller datasets.

Due to the well-established protocols and existing large database, this study relies on microsatellite markers. Nevertheless, in most comparative studies, SNPs performed equally better or outperformed microsatellite markers in relatedness and sibship analysis (Weinman et al. 2015; Flanagan and Jones 2019). With the very recent introduction of SNPs for *P. anguinus* (Recknagel et al. 2024a, Recknagel et al. 2024b), the implementation of both markers will soon be possible. The usage of SNPs will complete the results obtained by microsatellite markers and enable the comparison of methodology in the species. Selecting the best parameters and the best performing methodology will empower very precise and detailed determination of parentage and sibship among olms.

In the future, I suggest the complete analysis of all obtained samples - this will give insights into the family dispersion through the cave systems. The complete analysis will enable the composition of complex family relationships, thus giving a huge insight into the life of olms, and subterranean animals very hard to investigate. A very low recapture rate implies the existence of a much larger population, which is also an important consideration for parentage relationships. A lot of family members are not sampled, and many assembled families will have a large number of members missing

5.5. Population genetics and olm conservation

As suggested by Kostanjšek et al. (2023), olm acts as a flagship species in conservation attempts of the Dinaric karst subterranean habitats, indirectly supporting the protection of cave ecosystems and subterranean species. Implementation of population genetics data in the proteus population research enabled important and more detailed overview of the population genetic diversity and their connectivity within one large cave system. However, there are several challenges in the olm research, from obstacles in field work due to the habitat inaccessibility, to the complexity of its genome due to its size, are possible to overcome by the usage of microsatellites. Research on genetic diversity and population genetic structure represent an essential tool for effective conservation of concerned species, enabling faster inferences about population declines or recent bottlenecks, supporting management decisions in a shorter timeframe. Consequently, quicker response can be crucial for saving rapidly declining or endangered species (Storfer et al., 2014). Long term genetic data gives information about the changes in population, indirectly informing about the pollution of the habitat. Basic knowledge and monitoring of olm (*P. anguinus*) populations is necessary for the priority Habitats' directive species and further implementing of necessary conservation measures can indirectly help to conserve the entire groundwater ecosystem. This includes many smaller, unique and mainly small range and endemic subterranean organisms, from subterranean cnidarians (*Velkovrhia enigmatica*), tube worms (*Marifugia cavatica*), bivalve (*Congeria* spp.), springtails, beetles and spiders, to more obscure groups like palpigrades and pseudoscorpions, which are also part of the subterranean ecosystem's structure and energy flow. Protection of endangered species and their habitats is especially crucial in the frame of groundwater ecosystems, as the main source of the drinking water. Once damaged, these ecosystems can seldom be restored.

6. CONCLUSION

The study of the population genetic parameters of olms (*Proteus anguinus*) from the Postojna-Planina Cave System led to the following conclusions:

- A set of 22 microsatellite loci is useful for population genetic studies and identification of individual specimen for further mark-release-recapture studies which will enable estimation of population size.
- Population genetic parameters of olm populations from different sites in Postojna and Planina Caves showed comparable genetic diversity of populations on both sites of the system and lower genetic diversity in hydrologically separated Rak Channel.
- The genetic connectivity between the populations in Postojna and Planina is lower than shown in a previous study on a smaller sample size.
- The recapture of olms in one year interval was 14.1%. Most of the recaptured animals showed very high site fidelity and some long-distance migrations upstream and downstream were also observed.
- Most of the animals caught during this systematic survey were adults and only a very low number of juveniles were observed, indicating a more pronounced hiding behaviour of juveniles.
- The relatedness between the olms was the highest in Rak Channel.
- Full and half sibling relationships were suggested among olms captured in the Rak Channel, at the entrance and in the middle parts of Pivka Channel (Planina Cave). This indicates that a large microsatellite dataset available is sufficient to determine the sibship of olms in the Postojna-Planina Cave System.

REFERENCES

Anderson, A. D. and Weir, B. S., 2007. A maximum-likelihood method for the estimation of pairwise relatedness in structured populations. *Genetics*, 176 (1), 421–440.

Balázs, G., Lewarne, B. and Herczeg, G., 2015. *In situ* Underwater tagging of aquatic organisms: A test using the cave-dwelling olm, *Proteus anguinus*. *Annales Zoologici Fennici*, 52 (3), 160–166.

Balázs, G., Lewarne, B. and Herczeg, G., 2020. Extreme site fidelity of the olm (*Proteus anguinus*) revealed by a long-term capture–mark–recapture study. *Journal of Zoology*, 311 (2), 99–105.

Briegleb, W., 1962. Zur Biologie und Ökologie des Grottenolms (*Proteus anguinus* Laur. 1768). *Zeitschrift für Morphol. und Ökologie der Tiere*, 51, 271–334.

Brown, J. Z. and Johnson, J. E., 2001. Population biology and growth of Ozark cavefish in Logan Cave National Wildlife Refuge, Arkansas. *Environmental Biology of Fishes*, 62 (1–3), 161–169.

Chen, Z., Auler, A. S., Bakalowicz, M., Drew, D., Griger, F., Hartmann, J., Jiang, G., Moosdorf, N., Richts, A., Stevanovic, Z., Veni, G. and Goldscheider, N., 2017. The World Karst Aquifer Mapping project: concept, mapping procedure and map of Europe. *Hydrogeology Journal*, 25 (3), 771–785.

Deharveng, L., Bedos, A., Pipan, T. and Culver, D. C., 2024. Global subterranean biodiversity: A unique pattern. *Diversity*, 16 (3), 157.

Diez, C. M., Trujillo, I., Martinez-Urdiroz, N., Barranco, D., Rallo, L., Marfil, P. and Gaut, B. S., 2015. Olive domestication and diversification in the Mediterranean Basin. *New Phytologist*, 206 (1), 436–447.

Durand, J. P. and Delay, B., 1981. Influence of temperature on the development of *Proteus anguinus* (Caudata: Proteidae) and relation with its habitat in the subterranean world. *Journal of Thermal Biology*, 6 (1), 53–57.

Evanno, G., Regnaut, S. and Goudet, J., 2005. Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology*, 14 (8), 2611–2620.

Flanagan, S. P. and Jones, A. G., 2019. The future of parentage analysis: From microsatellites to SNPs and beyond. *Molecular Ecology*, 28 (3), 544–567.

Galpern, P., Manseau, M., Hettinga, P., Smith, K. and Wilson, P., 2012. Allelematch: an R package for identifying unique multilocus genotypes where genotyping error and missing data may be present. *Molecular Ecology Resources*, 12 (4), 771–778.

Gibert, J. and Deharveng, L., 2002. Subterranean ecosystems: A truncated functional biodiversity. *BioScience*, 52 (6), 473–481.

Goldscheider, N., Chen, Z., Auler, A. S., Bakalowicz, M., Broda, S., Drew, D., Hartmann, J., Jiang, G., Moosdorf, N., Stevanovic, Z. and Veni, G., 2020. Global distribution of carbonate rocks and karst water resources. *Hydrogeology Journal*, 28 (5), 1661–1677.

Gorički, Š. and Trontelj, P., 2006. Structure and evolution of the mitochondrial control region and flanking sequences in the European cave salamander Proteus anguinus. *Gene*, 378, 31–41.

Goudet, J., 2002. FSTAT Software. *Institute of Ecology BB, UNIL, CH-1015 (2002) Editor. 2.9.3.2 ed., Laussane.*

Gredar, T., Leonardi, A., Novak, M., Sepčić, K., Mali, L. B., Križaj, I. and Kostanjšek, R., 2019. Vitellogenin in the European cave salamander, *Proteus anguinus*: Its characterization and dynamics in a captive female as a basis for non-destructive sex identification. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 235, 30–37.

Guillame, O., Durand, J. and Juberthie, C., 1999. Spermatophores of *Proteus anguinus* (Caudata, Proteidae) and *Euproctus asper* (Caudata, Salamandridae). *Mémoires de Biospéologie*, 26, 25–28.

Guillaume, O., 2000. Role of chemical communication and behavioural interactions among conspecifics in the choice of shelters by the cave-dwelling salamander *Proteus anguinus* (Caudata, Proteidae). *Canadian Journal of Zoology*, 127, 263–272.

Gunn, J., 2004. *Encyclopedia of Caves and Karst Science*. Routledge.

Hamilton, M. B., 2009. *Population Genetics*. 1st ed. Chichester: A John Wiley & Sons Ltd.

Hauser, S. S., Athrey, G. and Leberg, P. L., 2021. Waste not, want not: Microsatellites remain an economical and informative technology for conservation genetics. *Ecology and Evolution*, 11 (22), 15800–15814.

Hohenlohe, P. A., Funk, W. C. and Rajora, O. P., 2021. Population genomics for wildlife conservation and management. *Molecular Ecology*, 30 (1), 62–82.

Jones, A. G. and Ardren, W. R., 2003. Methods of parentage analysis in natural populations. *Molecular Ecology*, 12 (10), 2511–2523.

Jones, A. G., Small, C. M., Paczolt, K. A. and Ratterman, N. L., 2010. A practical guide to methods of parentage analysis. *Molecular Ecology Resources*, 10 (1), 6–30.

Jones, O. R. and Wang, J., 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, 10 (3), 551–555.

Jones, W. K. and White, W. B., 2019. Karst. *In*: *Encyclopedia of Caves*. Elsevier, 609–618.

Juberthie, C., Durand, J. and Dupuy, M., 1996. La reproduction des Protées (*Proteus anguinus*): Bilan de 35 ans d'élevage dans les grottes-laboratoires de Moulis et d'Aulignac. *Mémoires de biospéologie*, 23, 53–56.

Kalinowski, S. T., Wagner, A. P. and Taper, M. L., 2006. ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes*, 6 (2), 576–579.

Koller Šarić, K., Jelić, D., Kovač Konrad, P., Jalžić, B., Aljančić, G., Sremac, J., Karaica, B., Bedek, J., Lukić, M., Lukač, M., Lewarne B., Balázc, G., Holtze, S., Legović, S., Božić, V., Bressi, N., Cizelj, I., Sket, B., Budić, M., Herczef, G., Braude, S., Göritz, F., Hermes, R., Berndt Hildebrandt, T., Mutschmann, F., Szentiks, C. A., Rnjak, G., Jalžić, V., Buzjak, N., Basara, D., Cvitanović, H., Ćukušić, I. and Polić, G., 2019. *Proteus*. Zagreb: Association Hyla.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. and Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, 15 (5), 1179–1191.

Kostanjšek, R., Zakšek, V., Bizjak-Mali, L. and Trontelj, P., 2023. The olm (*Proteus anguinus*), a flagship groundwater species. *In*: *Groundwater Ecology and Evolution*. Elsevier, 305–327.

Kretschmer, D., Wachholz, A. and Reinecke, R., 2023. Global groundwater in the Anthropocene. *In*: *Groundwater Ecology and Evolution*. Elsevier, 483–500.

Li, C. C., Weeks, D. E. and Chakravarti, A., 1993. Similarity of DNA fingerprints due to chance and relatedness. *Human Heredity*, 43 (1), 45–52.

Lunghi, E. and Bruni, G., 2018. Long-term reliability of visual Implant elastomers in the Italian cave salamander (*Hydromantes italicus*). *Salamandra*, 54 (4), 283–286.

Lunghi, E., Cianferoni, F., Corti, C., Zhao, Y., Manenti, R., Ficetola, G. F. and Mancinelli, G., 2022. The trophic niche of subterranean populations of *Speleomantes italicus*. *Scientific Reports*, 12 (1), 18257.

Lynch, M. and Ritland, K., 1999. Estimation of pairwise relatedness with molecular markers. *Genetics*, 152 (4), 1753–1766.

Lyu, D., Sun, S., Shan, X. and Wang, W., 2023. Inbreeding evaluation using microsatellite confirmed inbreeding depression in growth in the *Fenneropenaeus chinensis* natural population. *Frontiers in Genetics*, 14.

Malard, F., Griebler, C. and Rétaux, S., 2023. Groundwater ecology and evolution: an introduction. *In*: *Groundwater Ecology and Evolution*. Elsevier, xvii–xxix.

Mammola, S., Cardoso, P., Culver, D. C., Deharveng, L., Ferreira, R. L., Fišer, C., Galassi, D. M. P., Griebler, C., Halse, S., Humphreys, W. F., Isaia, M., Malard, F., Martinez, A., Moldovan, O. T., Niemiller, M. L., Pavlek, M., Reboleira, A. S. P. S., Souza-Silva, M., Teeling, E. C., Wynne, J. J. and Zagmajster, M., 2019. Scientists' warning on the conservation of subterranean ecosystems. *BioScience*, 69 (8), 641–650.

Manenti, R., Di Nicola, M. R., Zampieri, V., Grassi, G., Creanza, T., Mauri, E., Ficetola, G. F. and Barzaghi, B., 2024. Wandering outside of the Styx: Surface activity of an iconic subterranean vertebrate, the olm (*Proteus anguinus*). *Ecology*, 105 (3).

Marshall, T. C., Coltman, D. W., Pemberton, J. M., Slate, J., Spalton, J. A., Guinness, F. E., Smith, J. A., Pilkington, J. G. and Clutton–Brock, T. H., 2002. Estimating the prevalence of inbreeding from incomplete pedigrees. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269 (1500), 1533–1539.

Morin, P. A., Martien, K. K. and Taylor, B. L., 2009. Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*, 9 (1), 66–73.

Muñoz, I., Henriques, D., Jara, L., Johnston, J. S., Chávez-Galarza, J., De La Rúa, P. and Pinto, M. A., 2017. SNPs selected by information content outperform randomly selected microsatellite loci for delineating genetic identification and introgression in the endangered dark European honeybee (*Apis mellifera mellifera*). *Molecular Ecology Resources*, 17 (4), 783–795.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. and Shipley, P., 2004. micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4 (3), 535–538.

Pacheco, C., Rio-Maior, H., Nakamura, M., Álvares, F. and Godinho, R., 2024. Relatedness-based mate choice and female philopatry: inbreeding trends of wolf packs in a human-dominated landscape. *Heredity*, 132 (4), 211–220.

Patenković, A., Tanasković, M., Erić, P., Erić, K., Mihajlović, M., Stanisavljević, L. and Davidović, S., 2022. Urban ecosystem drives genetic diversity in feral honey bee. *Scientific Reports*, 12 (1), 17692.

Pathak, D. and Ali, S., 2012. Repetitive DNA: A tool to explore animal genomes/transcriptomes. *In*: *Functional Genomics*. InTech.

Peakall, R. and Smouse, P. E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28 (19), 2537–2539.

Pritchard, J. K., Stephens, M. and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155 (2), 945–959.

Puechmaille, S. J., 2016. The program Structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, 16 (3), 608–627.

Putman, A. I. and Carbone, I., 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, 4 (22), 4399–4428.

Queller, D. C. and Goodnight, K. F., 1989. Estimating relatedness using genetic markers. *Evolution*, 43 (2), 258–275.

Recknagel, H., Močivnik, L., Zakšek, V., Luo, Y., Kostanjšek, R. and Trontelj, P., 2024. Generation of genome-wide SNP markers from minimally invasive sampling in endangered animals and applications in species ecology and conservation. *Molecular Ecology Resources*, e13995.

Recknagel, H., Zakšek, V., Delić, T., Gorički, Š. and Trontelj, P., 2024. Multiple transitions between realms shape relict lineages of *Proteus* cave salamanders. *Molecular Ecology*, 33 (8), e16868.

Ritland, K., 1996. Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research*, 67 (2), 175–185.

Šebela, S., 2019. Postojna—Planina Cave System, Slovenia. *In*: *Encyclopedia of Caves*. Elsevier, 812–821.

Sket, B., 1997. Distribution of *Proteus* (Amphibia: Urodela: Proteidae) and its possible explanation. *Journal of Biogeography*, 24 (3), 263–280.

Sket, B., 2017. Discovering the black proteus *Proteus anguinus parkelj* (Amphibia: Caudata). *Natura Sloveniae*, 19 (1), 27–28.

Slatkin, M., 2008. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, 9 (6), 477–485.

Speed, D. and Balding, D. J., 2015. Relatedness in the post-genomic era: is it still useful? *Nature Reviews Genetics*, 16 (1), 33–44.

Storfer, A., Mech, S. G., Reudink, M. W. and Lew, K., 2014. Inbreeding and strong population subdivision in an endangered salamander. *Conservation Genetics*, 15 (1), 137–151.

Taylor, H. R., 2015. The use and abuse of genetic marker-based estimates of relatedness and inbreeding. *Ecology and Evolution*, 5 (15), 3140–3150.

Wang, J., 2002. An estimator for pairwise relatedness using molecular markers. *Genetics*, 160 (3), 1203–1215.

Wang, J., 2007. Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genetical Research*, 89 (3), 135–153.

Wang, J., 2011. Coancestry: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, 11 (1), 141–145.

Wang, J., 2017. The computer program Structure for assigning individuals to populations: easy to use but easier to misuse. *Molecular Ecology Resources*, 17 (5), 981–990.

Ward, R. D., 2006. The importance of identifying spatial population structure in restocking and stock enhancement programmes. *Fisheries Research*, 80 (1), 9–18.

Weinman, L. R., Solomon, J. W. and Rubenstein, D. R., 2015. A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources*, 15 (3), 502–511.

Willi, Y., Kristensen, T. N., Sgrò, C. M., Weeks, A. R., Ørsted, M. and Hoffmann, A. A., 2022. Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. *Proceedings of the National Academy of Sciences*, 119 (1).

Wright, S., 1931. Evolution in Mendelian populations. *Genetics*, 16 (2), 97–159.

Zagmajster, M., Malard, F., Eme, D. and Culver, D. C., 2018. Subterranean biodiversity patterns from global to regional scales. *In*: Moldovan, O., Kováč, L., and Halse, S., eds. *Cave Ecology*. Springer, Cham, 195–227.

Zagmajster, M., Polak, S. and Fišer, C., 2021. Postojna-Planina Cave System in Slovenia, a hotspot of subterranean biodiversity and a cradle of speleobiology. *Diversity*, 13 (6), 271.

Zakšek, V., Konec, M. and Trontelj, P., 2018. First microsatellite data on *Proteus anguinus* reveal weak genetic structure between the caves of Postojna and Planina. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 28 (1), 241–246.

Zakšek, V. and Trontelj, P., 2017. Conservation genetics of proteus in the Postojna-Planina Cave System. *Natura Sloveniae*, 19 (1), 33–34.

CURRICULUM VITAE

Dora Kermek was born on March 18th, 2001, in Čakovec, Croatia. She finished the First Gymnasium Varaždin as the best graduate of the school and the Varaždin County. She enrolled in the undergraduate study of molecular biology at the Faculty of Science, University of Zagreb, in 2019 and graduated in 2022 with the greatest honour (*summa cum laude*). In 2022, she started graduate studies in molecular biology at the same faculty. She spent third semester of graduate studies at Biotechnical faculty, Ljubljana for Erasmus+ mobility for studies. During her studies she participated in the work of several scientific laboratories: Laboratory of Cell Biophysics (Ruđer Bošković Institute, Zagreb), Laboratory of molecular genetics (Ruđer Bošković Institute, Zagreb), and Subterranean Biology Lab (Biotechnical faculty, University of Ljubljana). She is an active member of Croatian Biospeleological Society, Zagreb, and Biology Students Association – BIUS, Zagreb. She was the winner of the Rector's Award for individual scientific and artistic work (one or two authors) in the academic year 2021/22 for the scientific research: *Integrative approach in the research of stoneflies (Insecta: Plecoptera) and caddisflies (Insecta: Trichoptera) of Medvednica Nature Park with a comparative phylogeographical analysis of widely distributed species*.

List of publications:

Kermek, D., Pischiutta, N., Hlebec, D., Sivec, I., and Kučinić M., 2024. Utilising public sequence databases to investigate genetic diversity of stoneflies in Medvednica Nature Park. *Biodiversity Data Journal 12*: e121398.

Bedek, J., Šepčević, M., Kermek, D., Okićki, L., Kuharić, N., and Dražina, T., 2022. Biology Students Association as the base for establishing future experts and cave fauna exploration. In: *Proceedings of the 25th International Conference on Subterranean Biology*, 18 - 22 July 2022. Cluj-Napoca, Romania.

Kermek, D., Pischiutta, N., Hlebec, D., and Kučinić, M., 2022. Inventory and analysis of diversity for stoneflies (Plecoptera), caddisflies (Trichoptera) and scorpionflies (Mecoptera) in Medvednica Nature Park using classical taxonomy and DNA barcoding. In: *International Conference on DNA Barcoding and Biodiversity*, 25 - 27 May 2022. Sofia, Bulgaria. ISBN: 978-954-25-0382-8, pages 138 – 139.

Dražina, T., Šepčević, M., Kermek, D., Bedek, J., Čupić, I., Hlebec, D., Pavlek, M., Kuharić, N., Kirin, A., 2022. Recent biospeleological research in the Žumberak – Samobor hills Nature Park. In: *3rd Dinaric Symposium on Subterranean Biology*, 9 – 10 April 2022. Trebinje, Bosnia and Hercegovina. Page 23.

SUPPLEMENTARY MATERIAL

List of supplementary materials:

Figure S.1. Structure genetic cluster analysis of olms (*P. anguinus*) from Postojna (1, 2) and Planina (3, 4) caves into groups: $K = 2$ (A) and $K = 3$ (B) using 22 microsatellite loci without prior information on the location of individuals ($n = 288$).

Figure S.2. Structure genetic cluster analysis of olms (*P. anguinus*) from Postojna (1, 2) and Planina (3, 4) caves into groups: $K = 2$ (A) and $K = 3$ (B) using 22 microsatellite loci without prior information on the location of individuals ($n = 108$).

Figure S.3. Estimates of relatedness (r) among olms (*P. anguinus*) in Planina Cave (Pivka and Rak Channel) using triadic likelihood estimator

Table S.1. Summary information for the multiplexes (G1 – G4) and markers used for the amplification of 22 olm (*P. anguinus*) microsatellite loci.

Table S.2. Observed allelic frequencies at each microsatellite loci from two olm (*P. anguinus*) populations: Postojna Cave and Planina Cave.

Table S.3. List of the private alleles for each microsatellite loci of two olm (*P. anguinus*) populations: Postojna Cave and Planina Cave.

Table S.4. Observed allelic frequencies at each microsatellite loci of olms (*P. anguinus*) in Rak Channel.

Table S.5. Significant linkage disequilibrium among 22 used microsatellite loci in olms (*P. anguinus*), significance level < 0.01.

Table S.6. Relationships among olms (*P. anguinus*), with corresponding sample voucher code from analysed sections of Planina Cave using full-pedigree likelihood estimator (implemented in Colony) and maximum likelihood estimator (implemented in ML-Relate).

Table S.7. Likelihood ratio test for two *a priori* relationships between olm (*P. anguinus*) individuals with the corresponding sample voucher code obtained via ML-Relate program.

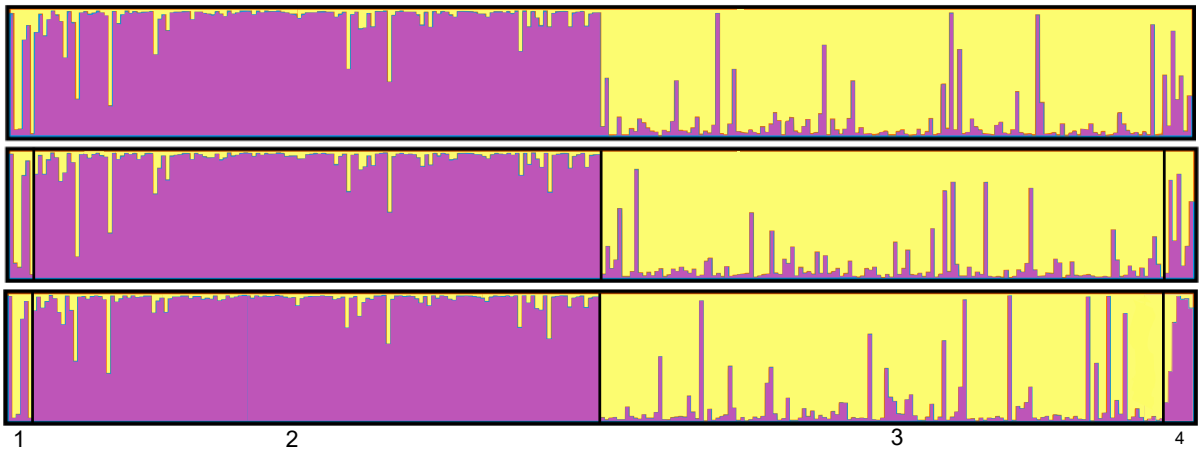

Figure S.1. Structure genetic cluster analysis of olms (*P. anguinus*) from Postojna (1, 2) and Planina (3, 4) caves into groups: $K = 2$ (A) and $K = 3$ (B) using 22 microsatellite loci without prior information on the location of individuals ($n = 288$). Each individual is represented by one column divided according to its probability of membership of "one population": violet population and yellow population (A), and additionally indigo population (B). First 144 samples represent samples from Postojna Cave (6 samples from Postojna (1) and 138 samples from Črna Cave (2)). Second 144 samples represent samples from Planina Cave, six samples randomly taken from each cave section. The order of Planina channel sections from left to right is A – Z (3), ending with Rak Channel (4). Each rectangle represents one out of three runs with different samples from Črna Cave and Pivka Channel.

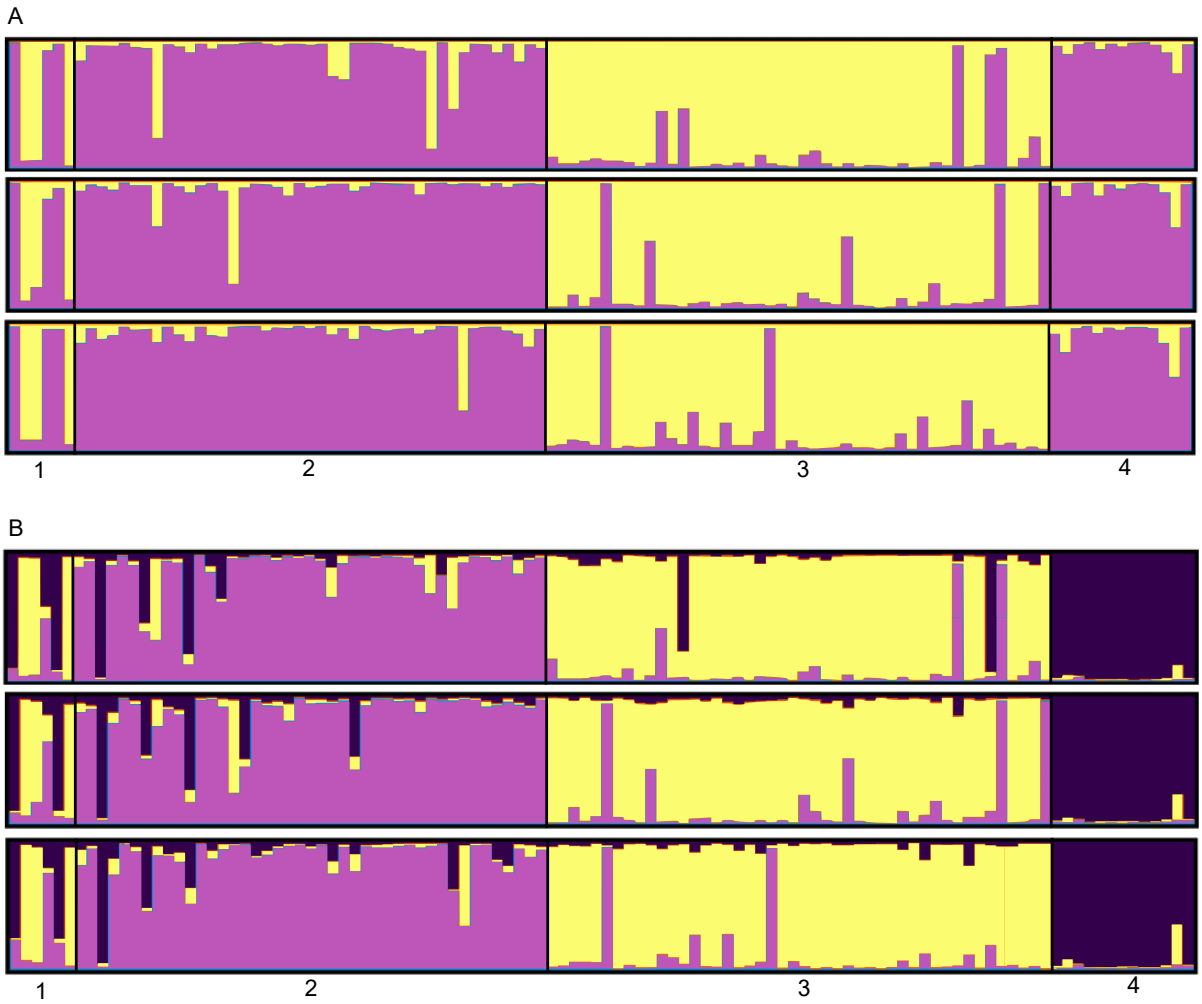

Figure S.2. Structure genetic cluster analysis of olms (*P. anguinus*) from Postojna (1, 2) and Planina (3, 4) caves into groups: $K = 2$ (A) and $K = 3$ (B) using 22 microsatellite loci without prior information on the location of individuals ($n = 108$). Each individual is represented by one column divided according to its probability of membership of "one population": violet population and yellow population (A), and additionally indigo population (B). First 49 samples represent samples from Postojna Cave (6 samples from Postojna (1) and 43 samples from Črna Cave (2)). Second 59 samples represent samples from Planina Cave (46 samples from Pivka (3) and 13 samples from Rak Channel (4)). The order of Planina channel sections from left to right is A – Z (3), ending with Rak Channel (4). Each rectangle represents one out of three runs with different samples from Črna Cave and Pivka Channel.

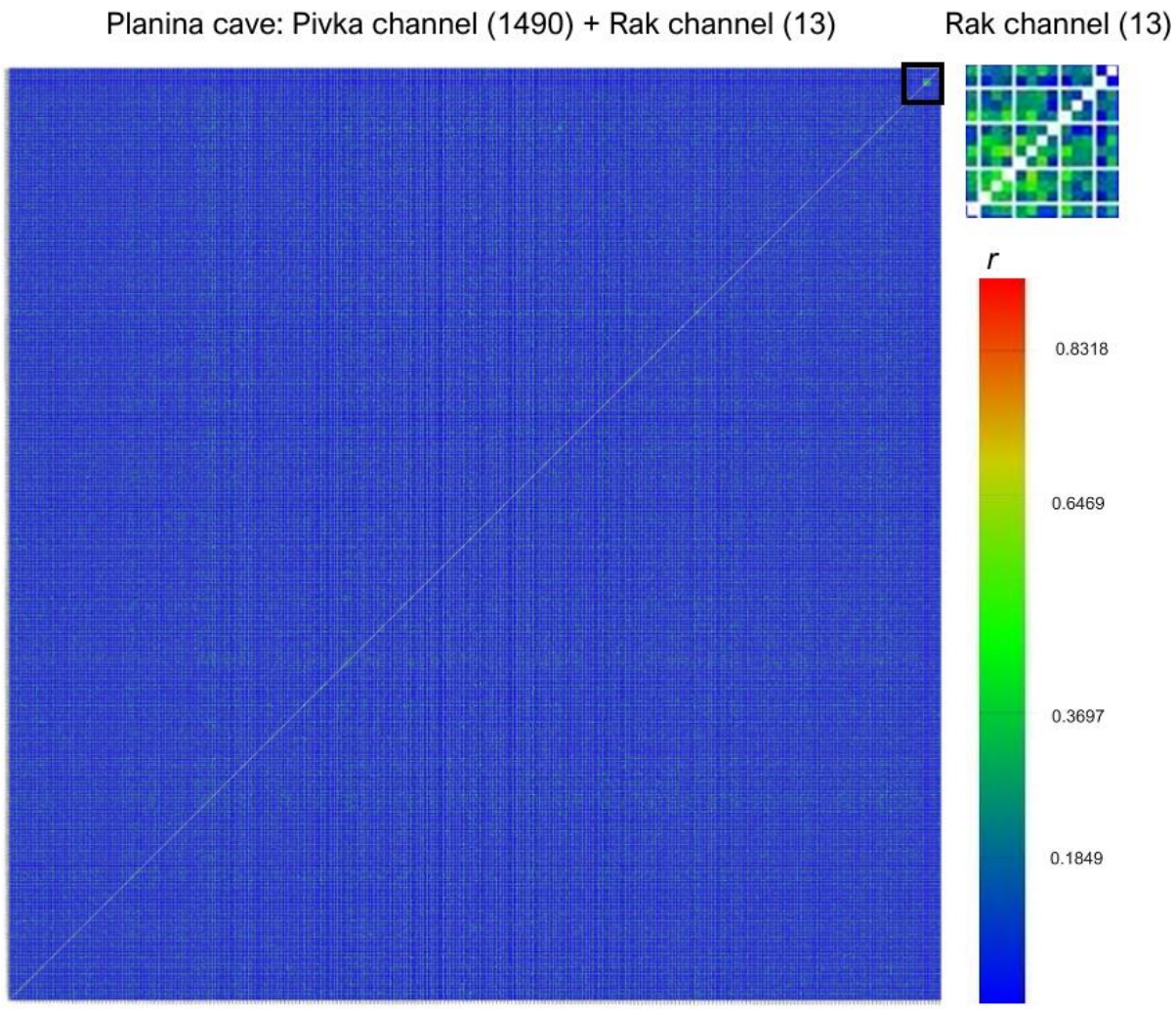

Figure S.3. Estimates of relatedness (*r*) among olms (*P. anguinus*) sampled from Planina Cave (Pivka and Rak Channel) using triadic likelihood estimator. Analysis included 1490 individuals from Pivka Channel and 13 individuals from Rak Channel. X and Y axis represent individuals, while each square on the plot represents the estimated relatedness value (e.g. the colour matching the legend) for given dyad, corresponding to individuals from X and Y axis. The data above diagonal line ($y = x$, in white) correlates with the data below the line. Relatedness among samples from Rak Channel is squared on the big plot and enlarged in the upper right corner of the picture

Table S.1. Summary information for the multiplexes (G1 – G4) and markers used for the amplification of 22 olm (*P. anguinus*) microsatellite loci. Annealing temperature is given for the whole multiplex. Loci without specified multiplex PCR were amplified in singleplex PCR. Abbreviations: abb. – abbreviation, Mod. – modification, T_a – annealing temperature.

Locus	**Abb.**	**Primer sequence 5'-3'**	**Repeat motif**	**Mod. (dye)**	**Length (bp)**	T_a / ^{O}C
		Multiplex G1				
PA_054846	PC11	F: TATCCTTCGAGCTCTGCACC R: CTACGTGCCACAGGTGTTTC	(ATAG)12	VIC	164–188	64
PA_069229	PC21	F: GCCAACAGGCTTGTACGTG R: TCCATCCACTTGGCTACGAC	(ATAG)14	NED	148–168	64
PA_254183	PC22	F: CAAGATGGGCCTTTGGTGC R: TCAAGGGGATGACTGCTGAC	(ATAG)13	NED	233–241	64
PA_005567	PC31	F: GTCCTATAGCGCCAAGATGC R: ACTACTGACACACAATAGCCAC	(ATAG)12	PET	168–176	64
PA_211285	PC01	F: AGAGCGCTTAACAAAACCCC R: AAGACTGCCCTACCTCAAGC	(GATA)12	FAM	122–160	64
		Multiplex G2				
PA_002648	PA01	F: TGCCGTAGTAGACTCTTGCC R: CGCAAGCACCAAGATAGCTC	(TCTA)22	FAM	156–176	64
PA_001534	PA22	F: ACGTTGCTTCATGTGCCTTC R: CCTTGGGTCGTATGTTGCAC	(ATCT)18	NED	197–241	64
PA_066527	PC12	F: GAAGGGCTGTGCTAATGCAG R: ACTGTTCCCCAGAATCCAGC	(TCTA)15	VIC	225–241	64
PA_004495	PC32	F: TTTCGCTAGTGCCTTCCATC R: AGAAAGACCATGCTCCCTGC	(GATA)10	PET	225–241	64
		Multiplex G3				
PA_024681	PA21	F: ATAGCTGTTCACGCAGAACG R: CCAGACTATATGCAGGCCCC	(AGAT)13	NED	135–143	62
PA_176651	PA31	F: TGACCTATTCGGTTGAACGTG R: AACGCTGTCATTTGCAGCTC	(GATA)17	PET	153–165	62
PA_237589	PB01	F: GAGGGCTAAACTAACATGCGG R: GGTGACCTTTGGGCTGTATC	(TCTA)23	FAM	143–181	62
PA_279833	PB03	F: TGAGGCTGCACACAGTTATG R: AGCAAACACTTGTGAGGGATG	(TCTA)20	FAM	290–319	62
		Multiplex G4				
PA_190694	PA32	F: TCCGCCTGTACCCATTTAGG R: TGACATGTTTTGAGCCACGG	(ATAG)19	PET	201–224	62
PA_028386	PB12	F: AGGTCTGCACAGCTAAAATAAATAC R: AGCATAAGCTATCGAAAACACAG	(GATA)20	VIC	330–357	62
PA_025902	PB21	F: GTAGGCATGCACAATGGGTC R: AGGCCTAGTTGGGTATCAGC	(ATAG)13	NED	120–154	62
PA_035455	PC03	F: AATGACGTTTGCCACCCAAG R: GTCTGAGCATCTTTAGCGCC	(TCTA)16	FAM	242–250	62

Table S.1. Summary information for the multiplexes (G1 – G4) and markers used for the amplification of 22 olm (*P. anguinus*) microsatellite loci. Annealing temperature is given for the whole multiplex. Loci without specified multiplex PCR were amplified in singleplex PCR. Abbreviations: abb. – abbreviation, Mod. – modification, T_a – annealing temperature.

Locus	**Abb.**	**Primer sequence 5'-3'**	**Repeat motif**	**Mod. (dye)**	**Length (bp)**	T_a / ^{O}C
Singleplex PCR						
PA_001895	PA02	F: ACTTCAGCCATCTCTCGGTC R: TCGCATTTCGCATGTACTCG	(TAGA)12	FAM	234–246	62
PA_111396	PA12	F: TGCCCAATAGGTGAAAAGCG R: CCCCACTCGGTAGTTGAGAC	(TAGA)15	VIC	203–215	62
PA_012650	PB02	F: GCCAAGATGGCTGTTCAGG R: AAGTACGACCACATCACCCC	(TAGA)18	FAM	211–230	60
PA_110343	PB22	F: ACCCAAGGGCATAGGAACTC R: ATGTGCGCTTAACAAATGGC	(CTAT)14	NED	227–244	60
PA_004734	PC02	F: AGCTCCATCGGTGAGATAGC R: CTCCCACAGCCTGAGAAGTC	(GATA)13	FAM	193–209	64

Table S.2. Observed allelic frequencies at each microsatellite loci from two olm (*P. anguinus*) populations: Postojna Cave and Planina Cave. Total number of samples: 1647, Postojna cave samples: 144, Planina cave samples: 1503. "n" indicates the number of successfully genotyped samples for each locus and population. Grey colour indicates rare alleles with the frequency less than 1%.

Locus	Allele/n	Postojna	Planina
PA01	**n**	144	1502
	152	/	0.00233
	156	0.17014	0.05326
	160	0.01736	0.00200
	164	0.00694	0.00300
	168	0.01389	0.01065
	172	0.78472	0.90280
	176	0.00694	0.02597
PA02	**n**	144	1500
	230	/	0.00100
	234	0.48264	0.53300
	238	0.01042	0.24400
	242	0.47917	0.20333
	246	0.02778	0.01867
PA12	**n**	144	1503
	199	0.00347	/
	203	0.14583	0.05822
	207	0.82292	0.85928
	211	0.02778	0.07418
	214	/	0.00732
	218	/	0.00033
	238	/	0.00067
PA21	**n**	144	1503
	134	0.11111	0.06055
	138	0.28819	0.55755
	143	0.60069	0.37758
	148	/	0.00432
PA22	**n**	144	1502
	181	0.01389	0.00133
	197	/	0.00133
	201	/	0.00799
	205	0.01389	0.04228
	209	0.50694	0.75466
	213	0.01042	0.03795
	221	0.02778	0.00100
	225	/	0.00033
	229	/	0.00300
	233	0.00694	0.00333
	237	0.38542	0.14148
	241	0.03472	0.00533
PA31	**n**	144	1503
	135	/	0.00033
	153	0.00347	0.00033
	157	0.17014	0.08150
	161	0.64236	0.81337
	165	0.18403	0.10379
	169	/	0.00067
PA32	**n**	144	1449
	197	/	0.00035
	201	0.01736	0.00311
	204	0.70486	0.29572
	208	0.02083	0.01829
	212	0.23958	0.17115
	216	/	0.03520
	220	0.01736	0.45790
	224	/	0.01829
PB01	**n**	144	1502
	138	/	0.00133
	142	0.03125	0.01332
	147	0.51042	0.47836
	152	0.01042	0.00766
	156	0.04861	0.00699
	160	/	0.00067
	164	0.03125	0.00866
	168	0.02778	0.10719
	172	/	0.01332
	176	0.32639	0.34421
	180	0.01389	0.01831
PB02	**n**	144	1502
	211	0.01389	0.00067
	215	0.02083	0.00533
	219	0.43403	0.16611
	223	0.50694	0.78795
	227	0.02083	0.03462
	230	0.00347	0.00533
PB03	**n**	144	1498
	265	/	0.00033
	281	/	0.00033
	285	/	0.00067
	289	0.01042	0.00300
	293	/	0.00167
	297	0.00694	0.02003
	302	0.00347	0.01736
	306	0.05903	0.24833
	311	0.82292	0.57310
	315	0.07639	0.08511
	319	0.02083	0.04706
	324	/	0.00300

Table S.2. Observed allelic frequencies at each microsatellite loci from two olm (*P. anguinus*) populations: Postojna Cave and Planina Cave. Total number of samples: 1647, Postojna cave samples: 144, Planina cave samples: 1503. "n" indicates the number of successfully genotyped samples for each locus and population. Grey colour indicates rare alleles with the frequency less than 1%.

Locus	Allele/n	Postojna	Planina	Locus	Allele/n	Postojna	Planina
PB12	**n**	143	1497	**PC03**	**n**	144	1503
	327	/	0.00167		**241**	0.19792	0.03360
	331	0.37762	0.18938		**245**	0.76042	0.93513
	335	0.02098	0.00868		**249**	0.04167	0.03127
	340	/	0.00234	**PC11**	**n**	144	1492
	344	0.00350	0.00200		**164**	0.07639	0.00804
	348	0.03497	0.04008		**168**	0.00347	0.00469
	351	0.29021	0.31797		**172**	0.90278	0.94806
	355	0.26923	0.40648		**176**	0.00347	0.01173
	358	0.00350	0.03140		**180**	/	0.00201
PB21	**n**	144	1486		**184**	0.01042	/
	120	/	0.00303		**188**	0.00347	0.02547
	124	/	0.00505	**PC12**	**n**	144	1500
	128	/	0.00034		**213**	/	0.00033
	132	0.00347	0.04744		**225**	0.00347	0.00333
	136	0.50694	0.42059		**229**	0.34375	0.16367
	140	0.28125	0.17194		**233**	0.23611	0.53267
	145	0.07986	0.04139		**237**	0.40625	0.29733
	149	0.12500	0.27591		**241**	0.01042	0.00267
	154	0.00347	0.03365	**PC21**	**n**	144	1503
	158	/	0.00067		**148**	0.01042	0.00699
PB22	**n**	144	1496		**152**	0.44792	0.46540
	228	0.00347	0.00234		**156**	0.00347	0.01497
	232	0.02778	0.00969		**160**	0.00347	0.00299
	236	0.91667	0.75902		**164**	0.48264	0.50931
	240	0.03472	0.03710		**168**	0.05208	0.00033
	244	0.01736	0.19184	**PC22**	**n**	144	1491
PC01	**n**	144	1503		**229**	0.00347	/
	122	/	0.00299		**233**	0.37500	0.18243
	125	/	0.00067		**237**	0.02778	0.26626
	130	/	0.00067		**241**	0.59375	0.54930
	134	0.47222	0.36494		**245**	/	0.00201
	152	0.10069	0.07152	**PC31**	**n**	144	1503
	156	0.42014	0.55489		**164**	/	0.00466
	160	0.00694	0.00399		**168**	0.02431	0.00499
	164	/	0.00033		**172**	0.56944	0.73320
PC02	**n**	144	1499		**176**	0.40278	0.25615
	193	0.04861	0.13776		**180**	0.00347	0.00100
	197	0.02083	0.10640	**PC32**	**n**	144	1495
	201	0.90972	0.73549		**225**	/	0.01204
	205	0.01389	0.01835		**229**	0.84375	0.74783
	209	/	0.00167		**233**	0.00694	0.00468
	213	0.00694	/		**237**	0.14931	0.23344
	221	/	0.00033		**241**	/	0.00201

Table S.3. List of the private alleles for each microsatellite loci of two olm (*P. anguinus*) populations: Postojna Cave and Planina Cave. Total number of samples: 1647, Postojna cave samples: 144, Planina cave samples: 1503. Postojna cave starts with "*" symbol for quicker and easier recognition. Yellow colour indicates alleles that are present in only one individual.

Locus	Population	Allele	Freq	Locus	Population	Allele	Freq
PA01	Planina cave	152	0.0023	*PB12*	Planina cave	327	0.0017
PA02	Planina cave	230	0.0010			340	0.0023
PA12	*Postojna cave	199	0.0035	*PB21*	Planina cave	128	0.0003
	Planina cave	218	0.0003			158	0.0007
		238	0.0007			120	0.0030
		214	0.0073			124	0.0050
PA21	Planina cave	148	0.0043	*PC01*	Planina cave	164	0.0003
PA22	Planina cave	225	0.0003			125	0.0007
		197	0.0013			130	0.0007
		229	0.0030			122	0.0030
		201	0.0080	*PC02*	*Postojna cave	213	0.0069
PA31	Planina cave	135	0.0003		Planina cave	221	0.0003
		169	0.0007			209	0.0017
PA32	Planina cave	197	0.0003	*PC11*	*Postojna cave	184	0.0104
		224	0.0183		Planina cave	180	0.0020
		216	0.0352	*PC12*	Planina cave	213	0.0003
PB01	Planina cave	160	0.0007	*PC22*	*Postojna cave	229	0.0035
		138	0.0013		Planina cave	245	0.0020
		172	0.0133	*PC31*	Planina cave	164	0.0047
PB03	Planina cave	265	0.0003	*PC32*	Planina cave	241	0.0020
		281	0.0003			225	0.0120
		285	0.0007				
		293	0.0017				
		324	0.0030				

Table S.4. Observed allelic frequencies at each microsatellite loci of olms (*P. anguinus*) in Rak Channel. Total number of samples: 13. This group of samples has no rare alleles with the frequency less than 1%.

Locus	Allele	Rak Channel	Locus	Allele	Rak Channel
PA01	**152**	0.002	**PB12**	**331**	0.231
	156	0.167		**335**	0.038
	168	0.042		**340**	0.231
	172	0.75		**351**	0.385
PA02	**234**	0.192		**355**	0.114
	238	0.154	**PB21**	**140**	0.385
	242	0.654		**149**	0.5
PA12	**203**	0.038		**154**	0.038
	207	0.885		**158**	0.077
	211	0.038	**PB22**	**228**	0.115
	218	0.038		**236**	0.731
PA21	**134**	0.269		**240**	0.077
	138	0.308		**244**	0.077
	143	0.385	**PC01**	**134**	0.769
	148	0.038		**152**	0.038
PA22	**209**	1		**156**	0.192
PA31	**157**	0.308	**PC02**	**193**	0.115
	161	0.692		**201**	0.846
PA32	**204**	0.462		**221**	0.038
	208	0.231	**PC03**	**245**	1
	212	0.308	**PC11**	**164**	0.308
PB01	**147**	0.0385		**172**	0.538
	152	0.038		**188**	0.154
	164	0.077	**PC12**	**229**	0.577
	168	0.154		**233**	0.231
	176	0.346		**237**	0.192
PB02	**219**	0.038	**PC21**	**156**	0.269
	223	0.923		**164**	0.731
	230	0.038	**PC22**	**233**	0.346
PB03	**281**	0.038		**241**	0.615
	306	0.077		**245**	0.038
	311	0.5	**PC31**	**168**	0.115
	315	0.077		**172**	0.115
	319	0.308		**176**	0.769
			PC32	**229**	0.846
				237	0.154

Table S.5. Significant linkage disequilibrium among 22 used microsatellite loci in olms (*P. anguinus*), significance level < 0.01. Numbers refer to the number of linked loci with specific locus. Shade of grey corresponds to the number size, white being the smallest (no linked loci) and dark grey being the greatest. Abbreviations: Post. – Postojna cave, Plan – Planina cave, Rak – Rak channel, *NLP* – total number of linked loci pairs, * - dataset with Rak channel individuals excluded from Planina cave individuals. Sizes of datasets: Postojna cave 144 individuals, Planina cave 1503 individuals; Postojna cave* 144 individuals, Planina cave* 1490 individuals (without Rak channel samples), Rak channel* 13 individuals.

Locus	*PA01*	*PA02*	*PA12*	*PA21*	*PA22*	*PA31*	*PA32*	*PB01*	*PB02*	*PB03*	*PB12*	
Post.	1	3	2	1	3	1	3	2	0	1	0	
Plan.	4	4	1	4	2	2	7	5	1	0	7	
*Post.	1	3	2	1	3	1	3	2	0	1	0	
*Plan.	2	3	3	4	3	2	6	4	1	1	5	
*Rak	1	0	0	1	0	0	0	1		0	0	
Locus	*PB21*	*PB22*	*PC01*	*PC02*	*PC03*	*PC11*	*PC12*	*PC21*	*PC22*	*PC31*	*PC32*	*NLP*
Post.	1	0	0	3	2	0	2	3	0	0	0	14
Plan.	11	3	2	1	1	8	2	5	6	5	3	42
*Post.	1	0	0	3	2	0	2	3	0	0	0	14
*Plan.	7	3	1	0	2	6	1	0	5	2	3	32
*Rak	0	1	0	0	0	0	0	0	0	0	0	2

Figure S.6. Relationships among olms (*P. anguinus*), with corresponding sample voucher code from analysed sections of Planina Cave using full-pedigree likelihood estimator (implemented in Colony) and maximum likelihood estimator (implemented in ML-Relate). Triplicate runs were performed for full-pedigree likelihood estimator, with obtained probabilities for each pair included in the table. LnL(R) represents natural logarithm of R, e.g. the relationship with the highest likelihood. In case the most probable relationship in ML-Relate differed from the relationship obtained by Colony, the abbreviation of that relationship is written instead of *LnL (R):* PO – parent-offspring, HS – half-sibling, U – unrelated. FS* means that full sibling relationships was confirmed after the specific hypothesis test. Control full sibling pair from Cave Laboratory Tular (PC522, PC523) has (K) next to their voucher code.

Cave section	***Offspring ID1***	***Offspring ID2***	***Full-pedigree likelihood***			***Maximum likelihood***
			Prob. 1	***Prob. 2***	***Prob. 3***	***LnL(R)***
Entrance sections of Planina cave			Full sibling pairs			
	PA365	PC844	0.999	0.999	0.999	-43.85
	PA362	PC650	0.998	0.998	0.998	-51.8
	PC848	PC994	0.994	0.994	0.994	-35.82
	PA585	PB617	0.993	0.993	0.993	-40.99
	PC522 (K)	PC523 (K)	0.984	0.984	0.984	-76.55
	PA361	PA362	0.975	0.975	0.975	-54.33
	PA361	PC650	0.947	0.947	0.947	-49.18
			Half sibling pairs			
	PA365	PC844	0.999	0.999	0.999	-43.85
	PA362	PC650	0.998	0.998	0.998	-51.8
	PC848	PC994	0.994	0.994	0.994	-35.82
	PA585	PB617	0.993	0.993	0.993	-40.99
	PC522 (K)	PC523 (K)	0.984	0.984	0.984	-76.55
	PA361	PA362	0.975	0.975	0.975	-54.33
	PA361	PC650	0.947	0.947	0.947	-49.18
L section of Pivka channel			Full sibling pairs			
	PC522 (K)	PC523 (K)	0.984	0.984	0.984	-82.9
	PB083	PB090	0.967	0.967	0.967	-57.01
	PB178	PC000	0.847	0.847	0.847	FS*
	PB983	PC000	0.736	0.736	0.736	-53.65
	PB178	PB983	0.491	0.491	0.491	-53.63
	PB189	PB190	0.45	0.45	0.45	PO
	PB093	PB994	0.389	0.389	0.389	PO
			Half sibling pairs			
	PB083	PB099	0.695	0.695	0.695	-60.63
	PB083	PB177	0.669	0.669	0.669	-64.52
	PB090	PB099	0.649	0.649	0.649	-57.13
	PB090	PB177	0.606	0.606	0.606	-60.79

Figure S.6. Relationships among olms (*P. anguinus*), with corresponding sample voucher code from analysed sections of Planina Cave using full-pedigree likelihood estimator (implemented in Colony) and maximum likelihood estimator (implemented in ML-Relate). Triplicate runs were performed for full-pedigree likelihood estimator, with obtained probabilities for each pair included in the table. LnL(R) represents natural logarithm of R, e.g. the relationship with the highest likelihood. In case the most probable relationship in ML-Relate differed from the relationship obtained by Colony, the abbreviation of that relationship is written instead of *LnL (R):* PO – parent-offspring, HS – half-sibling, U – unrelated. FS* means that full sibling relationships was confirmed after the specific hypothesis test. Control full sibling pair from Cave Laboratory Tular (PC522, PC523) has (K) next to their voucher code.

Cave section	***Offspring ID1***	***Offspring ID2***	***Full-pedigree likelihood***			***Maximum likelihood***
			Prob. 1	***Prob. 2***	***Prob. 3***	***LnL(R)***
Rak channel	Full sibling pairs					
	PB256	PB258	0.987	0.986	0.986	PO
	PB256	PB261	0.970	0.974	0.971	PO
	PC522 (K)	PC523 (K)	0.963	0.963	0.963	-52.36
	PB257	PB259	0.625	0.628	0.632	PO
	PB258	PB261	0.375	0.376	0.375	HS
	Half sibling pairs					
	PB265	PC522 (K)	0.978	0.977	0.977	U
	PB260	PB261	0.947	0.947	0.947	U
	PB254	PB265	0.931	0.932	0.926	U
	PB259	PC522 (K)	0.924	0.925	0.921	U
	PB253	PB262	0.902	0.902	0.902	U
	PB254	PC523 (K)	0.897	0.898	0.892	U
	PB254	PC522 (K)	0.866	0.867	0.861	U
	PB258	PB262	0.859	0.859	0.858	U
	PB259	PC523 (K)	0.845	0.845	0.842	U
	PB265	PC523 (K)	0.816	0.816	0.815	U
	PB256	PB262	0.815	0.815	0.814	U
	PB256	PB260	0.814	0.817	0.814	U
	PB258	PB260	0.808	0.810	0.808	U
	PB261	PB262	0.784	0.784	0.784	U
	PB257	PC522 (K)	0.722	0.722	0.722	HS
	PB255	PB260	0.646	0.650	0.647	U
	PB257	PC523 (K)	0.579	0.579	0.579	U

Table S.7. Likelihood ratio test for two *a priori* relationships between olm (*P. anguinus*) individuals with the corresponding sample voucher code obtained via ML-Relate program. $p < 0.05$ of specific hypothesis test indicates that putative relationship fits the data significantly better than the alternative relationship. Significantly better relationships are marked in grey

Cave section	***Offspring ID1***	***Offspring ID2***	***Putative relationship***	***Alternative relationship***	***p-value***
Rak channel	PB258	PB256	Parent-offspring	Full siblings	0.2511
			Full siblings	Parent-offspring	0.11059
	PB261	PB256	Parent-offspring	Full siblings	0.18476
			Full siblings	Parent-offspring	0.19114
	PB261	PB258	Parent-offspring	Full siblings	0.00896
			Full siblings	Parent-offspring	0.82893
	PB259	PB257	Parent-offspring	Full siblings	0.07385
			Full siblings	Parent-offspring	0.4384
Entrance	PC844	PA562	Parent-offspring	Full siblings	0.153
			Full siblings	Parent-offspring	0.04016
L section	PB178	PC000	Parent-offspring	Full siblings	0.17645
			Full siblings	Parent-offspring	0.19417
	PB189	PB190	Parent-offspring	Full siblings	0.09001
			Full siblings	Parent-offspring	0.37075